

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:39:13 ; Search time 11.1 Seconds  
(without alignments)  
788.344 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 1185  
Sequence: 1 MGPLPRTVELFYDLSPYSW.....AHLGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1180	99.6	225	1 GTK1_HUMAN	Q9Y2Q3 homo sapien
2	869	73.3	225	1 GTK1_RAT	P24473 rattus norv
3	328.5	27.7	226	1 YS21_CAEEL	Q09652 caenorhabdi
4	298.5	25.2	225	1 YS17_CAEEL	Q18973 caenorhabdi
5	95	8.0	203	1 NAHD_PSEPU	Q51948 pseudomonas
6	89.5	7.6	962	1 PTRA_ECOLI	P05458 escherichia
7	89	7.5	212	1 NAHD_PSESP	Q52462 pseudomonas
8	89	7.5	592	1 THD1_ARATH	Q9ZSS6 arabidopsis
9	88.5	7.5	613	1 SG2_BOVIN	P20616 bos taurus
10	84.5	7.1	316	1 MIAA_TREPA	O83644 treponema p
11	84.5	7.1	971	1 S3B1_YEAST	P49955 saccharomyc
12	83.5	7.0	488	1 CATA_LISSE	P24168 listeria se
13	83.5	7.0	700	1 NCD_DROME	P20480 drosophila
14	81.5	6.9	619	1 SG2_RAT	P10362 rattus norv
15	81.5	6.9	1061	1 DPOL_ADE12	P06538 human adeno
16	81	6.8	274	1 YG73_SYNY3	P74261 synchocyst
17	81	6.8	966	1 ST10_MOUSE	O55098 mus musculu
18	80.5	6.8	705	1 HSER_PIG	P32562 saccharomyc
19	80.5	6.8	1073	1 CS5_YEAST	P55204 sus scrofa
20	80	6.8	540	1 TOP1_AOUAE	O68893 aquifex aeo
21	80	6.8	1162	1 BXEN_CLOBO	P46082 clostridium
22	79.5	6.7	547	1 CATX_BAGSU	P94377 bacillus su
23	79.5	6.7	1073	1 HSER_HUMAN	P25092 homo sapien
24	78.5	6.6	383	1 PNAI_RICPR	P41077 rickettsia
25	78	6.6	1162	1 BXEN_CLOBO	Q06366 clostridium
26	77.5	6.5	634	1 YC36_METUA	O58633 methanococc
27	77	6.5	494	1 ADRO_RAT	P56322 rattus norv
28	76.5	6.5	962	1 YBX7_SCHPO	Q10201 schizosacch
29	76.5	6.5	1056	1 DPOL_ADE02	P03261 human adeno
30	76.5	6.5	1193	1 DPOL_ADE04	P87503 human adeno
31	76	6.4	446	1 PKB3_HUMAN	P15120 homo sapien
32	76	6.4	646	1 NODQ_RHIS3	P72339 r nodq bifu
33	76	6.4	2208	1 POLN_WANCV	Q69014 manchester

34	75.5	6.4	744	1 GYRA_AOUAE	O67108 aquifex aeo
35	75.5	6.4	887	1 GLND_KLEPN	P41393 klebsiella
36	75.5	6.4	1056	1 DPOL_ADE05	P04495 human adeno
37	75.5	6.4	1398	1 PLS_PYRFU	P72186 pyrococcus
38	75.5	6.4	1434	1 RPOB_UREPA	Q9PQV6 ureaplasma
39	75	6.3	231	1 LOLD_NEIMA	P57030 neisseria m
40	75	6.3	393	1 HEMX_ECOLI	P09127 escherichia
41	75	6.3	593	1 KPYA_TOBAC	O40545 nicotiana t
42	75	6.3	631	1 OE66_NPVLS	P89518 leucania se
43	75	6.3	686	1 MEPD_PIG	P47788 sus scrofa
44	75	6.3	925	1 PIPL_YEAST	P40020 saccharomyc
45	75	6.3	984	1 SECA_AOUAE	O67718 aquifex aeo

ALIGNMENTS

RESULT	1	GTCL_HUMAN	STANDARD:	PRT:	225 AA.
ID	GTCL_HUMAN	STANDARD:	PRT:	225 AA.	
AC	Q9Y2Q3: Q9PIS4:				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DE	Glutathione S-transferase, mitochondrial (EC 2.5.1.18) (GST 13-13)				
DE	Glutathione S-transferase subunit 13) (GST class-kappa) (HDCMD47P).				
GN	GSTK1				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Blood;				
RX	MEDLINE=20499367: PubMed=11042152;				
RA	Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,				
RA	Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,				
RA	Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.,				
RT	"Cloning and functional analysis of cDNAs with open reading frames for				
RT	300 previously undefined genes expressed in CD34+ hematopoietic				
RT	stem/progenitor cells."				
RT	Genome Res. 10:1546-1560(2000).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RA	Zhao Z., Huang X., Li N., Zhu X., Cao X.;				
RT	"A novel gene from human dendritic cell."				
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Uterus;				
RX	MEDLINE=21154917: PubMed=11230166;				
RA	Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,				
RA	Ansorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,				
RA	Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,				
RA	Mewes H.-W., Ottenwaelter B., Obermaier B., Tampe J., Haubner D.,				
RA	Wamburt R., Korn B., Klein M., Poustka A.;				
RT	"Towards a catalog of human genes and proteins: sequencing and				
RT	analysis of 500 novel complete protein coding human cDNAs.";				
RT	Genome Res. 11:422-435(2001).				
CC	-!- FUNCTION: MIGHT CONFER PROTECTION AGAINST GENOTOXIC AND CYTOTOXIC				
CC	ELECTROPHILES IN THE MITOCHONDRIAL COMPARTMENT (BY SIMILARITY).				
CC	-!- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.				
CC	-!- SUBUNIT: HOMODIMER (BY SIMILARITY).				
CC	-!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).				
CC	-!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/				
CC	or send an email to license@isb-sib.ch).				

CC EMBL; AF070657; AAD20963.1; -  
 DR EMBL; AF068287; AAF65506.1; -  
 DR EMBL; AL136938; CAB66872.1; -  
 DR MIM; 602321; -  
 DR InterPro; IPR004287; HCCA\_isomerase.  
 DR Pfam; PF03046; HCCA\_isomerase; 1.  
 KW Transferase; Mitochondrion.  
 FT INIT\_MET 0 BY SIMILARITY.  
 FT CONFLICT 178 178 G -> R (IN REF. 2).  
 FT CONFLICT 219 219 P -> S (IN REF. 2).  
 SQ SEQUENCE 225 AA; 25365 MW; FE91A5E0F0B0BA1 CRC64;

Query Match 99.6%; Score 1180; DB 1; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-95;  
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPLPRTVELFDVLSYSLGFEILCRYQINWININLQRLPSLITGIMKDSGNKPPGLLPR 61  
 DB 1 GPLPRTVELFDVLSYSLGFEILCRYQINWININLQRLPSLITGIMKDSGNKPPGLLPR 60  
 QY 62 KGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121  
 DB 61 KGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 120  
 QY 122 WMRVWSNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 181  
 DB 121 WMRVWSNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 180  
 QY 182 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 226  
 DB 181 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 225

RESULT 2  
 ID GTK1\_RAT STANDARD; PRT; 225 AA.  
 AC P24473; O09034;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glutathione S-transferase, mitochondrial (EC 2.5.1.18) (GST 13-13)  
 GN GSTK1 OR GSTK1-1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=97079244; PubMed=8920976;  
 RA Pemble S.E., Wardle A.F., Taylor J.B.;  
 RT "Glutathione S-transferase class kappa: characterization by the  
 RT cloning of rat mitochondrial GST and identification of a human  
 RT homologue.";  
 RL Biochem. J. 319:749-754 (1996).  
 RN [2]  
 RP SEQUENCE OF 1-33.  
 RX MEDLINE=91354194; PubMed=1883325;  
 RA Harris M.J., Meyer D.J., Coles B., Ketterer B.;  
 RT "A novel glutathione transferase (13-13) isolated from the matrix of  
 RT rat liver mitochondria having structural similarity to class theta  
 RT enzymes.";  
 RL Biochem. J. 278:137-141 (1991).  
 CC -!- FUNCTION: MIGHT CONFER PROTECTION AGAINST GENOTOXIC AND CYTOTOXIC  
 CC ELECTROPHILES IN THE MITOCHONDRIAL COMPARTMENT.  
 CC -!- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.  
 CC -!- SUBUNIT: HOMODIMER.  
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
 or send an email to license@isb-sib.ch).  
 DR EMBL; S83436; AAB50831.1; -  
 DR PIR; S17164; S17164.  
 DR InterPro; IPR004287; HCCA\_isomerase.  
 DR Pfam; PF03046; HCCA\_isomerase; 1.  
 KW Transferase; Mitochondrion.  
 FT INIT\_MET 0  
 FT CONFLICT 1 1 G -> C (IN REF. 2).  
 SQ SEQUENCE 225 AA; 25362 MW; FC895B730655E0C9 CRC64;

Query Match 73.3%; Score 869; DB 1; Length 225;  
 Best Local Similarity 69.3%; Pred. No. 4.7e-58;  
 Matches 156; Conservative 38; Mismatches 31; Indels 0; Gaps 0;

QY 2 GPLPRTVELFDVLSYSLGFEILCRYQINWININLQRLPSLITGIMKDSGNKPPGLLPR 61  
 DB 1 GPLPRTVELFDVLSYSLGFEILCRYQINWININLQRLPSLITGIMKDSGNKPPGLLPR 60  
 QY 62 KGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121  
 DB 61 KGLYMANDLKLRLHHLQIPHFPPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 120  
 QY 122 WMRVWSNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 181  
 DB 121 WMRVWSNEDITEPQSILAAAEKAGMSAQOGLLEKIATPKVKNQKLTETEAACRYGAF 180  
 QY 182 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 226  
 DB 181 GLPITVAHVQDQTHLMFGSDRMELLAHLLGKWMGPIPPAVNARL 225

RESULT 3  
 ID YS21\_CAEEL STANDARD; PRT; 226 AA.  
 AC Q09652;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 26.5 kDa protein 2K1320.1 in chromosome II.  
 GN 2K1320.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Berks M.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announcement/  
 or send an email to license@isb-sib.ch).  
 DR EMBL; Z46934; CAA87039.1; -  
 DR WormPep; ZK1320.1; CE01698.  
 DR InterPro; IPR004287; HCCA\_isomerase.  
 DR Pfam; PF03046; HCCA\_isomerase; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 226 AA; 26482 MW; DBA39A5994300164 CRC64;

us-09-441-723-1-rsp

Tue May 21 08:04:02 2002

120 GLWHRFYAYGKPIFTKSQ---VAEVLRLDLHVQVDELVMMSDSAEVKNILRENTDEAGN 176

179 GAFGLP---ITVAHVDGQT-HMLFGSDRMELLALHLLGKWMGPI 218  
177 GCGFAGPWHHTDGH--GKVLQTVFGSDRLPOVADFLAEPPFKGPM 218

RESULT 5

NAHD\_PSEPU STANDARD; PRT; 203 AA.  
AC Q51948;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 2-hydroxychromene-2-carboxylate isomerase (HCCA isomerase).  
GN NAHD.  
OS Pseudomonas putida.  
OG Plasmid NAH7.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=G7 / ATCC 17485;  
RX MEDLINE=95095951; PubMed=8002605;  
RA Eaton R.W.;  
RT "Organization and evolution of naphthalene catabolic pathways:  
sequence of the DNA encoding 2-hydroxychromene-2-carboxylate  
isomerase and trans-o-hydroxybenzylidenepyruvate hydratase-aldolase  
from the NAH7 plasmid.";  
RL J. Bacteriol. 176:7757-7762(1994).  
RN [2]  
RP CHARACTERIZATION.  
RC STRAIN=G7 / ATCC 17485;  
RX MEDLINE=93077433; PubMed=1447127;  
RA Eaton R.W., Chapman P.J.;  
RT "Bacterial metabolism of naphthalene: construction and use of  
recombinant bacteria to study ring cleavage of 1,2-  
dihydroxynaphthalene and subsequent reactions.";  
RT J. Bacteriol. 174:7542-7554(1992).  
RL -1- FUNCTION: CATALYZES THE ISOMERIZATION OF 2-HYDROXYCHROMENE-2-  
CARBOXYLATE (HCCA) TO TRANS-O-HYDROXYBENZYLIDENEPYRUVATE (THPPA).  
CC THE OPTIMUM PH FOR THE ENZYME IS 10. THE REACTION IS REVERSIBLE.  
CC -1- COFACTOR: GLUTATHIONE SEEMS TO STABILIZE THE ENZYME, WHICH LOSES  
ACTIVITY RAPIDLY IN THE ABSENCE OF THIS COMPOUND.  
CC -1- PATHWAY: UPPER NAPHTHALENE CATABOLIC PATHWAY WHICH INVOLVES  
CONVERSION OF NAPHTHALENE TO SALICYLATE.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U09057; AAA66358.1;  
DR InterPro: IPR004287; HCCA\_isomerase.  
DR Pfam; PF03046; HCCA\_isomerase; 1.  
KW Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
SQ SEQUENCE 203 AA; 23061 MW; E84B56F21C604945 CRC64;

Query Match 8.0%; Score 95; DB 1; Length 203;  
Best Local Similarity 24.2%; Pred. No. 0.28;  
Matches 51; Conservative 35; Mismatches 97; Indels 28; Gaps 8;

QY 8 VELYDVLSPYSWLGFEILCRYQNIWNINLQRLPSLTITGIMKDSGNKPPGL--LPRKGLY 65  
DB 3 VDFYDFLSPFYSYLANQRSLKLAQDYGLTIRYNAIDARVIAICNVGSPNRDLKVKLDY 62  
QY 66 MANDLKLRLHLLQIPHFPKDFLSVMLEK---SLSAMRFTAVNLEHPEMLEKASRELW 122

Query Match 27.7%; Score 328.5; DB 1; Length 226;  
Best Local Similarity 33.0%; Pred. No. 2.2e-21;  
Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;

QY 1 MGPLPRTVELFVDVLSWLGFEILCRYQNIWNINLQRLPSLTITGIMKDSGNKPPGL 59  
DB 1 MPKLPR-IDFYDVLSPYSYIAFEVQKLEMTQKGVITIRYPPFLGAVMKESGNRPAML 59  
QY 60 PRKGLYANDLKLRLHLLQIPHFPKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
DB 60 PARSIMMTDLRTAKTFWDIPPLPFLFWIWKYRTGAMKVLVLLVLEQDKELMLRAAR 119  
QY 120 ELWMRVWSRNEDITEPQSILAAEKAGMSAEQAQGLLEKIATPKVKNLKETTEACRYG 179  
DB 120 EMVRLWSRSEKIFEDDFEVLKAVGV--KNPEQIVKESKDEKIYKILMENTNKGVDLM 177  
QY 180 AFGLP-ITVAHVDGQTHMLFGSDRMELLALHLLGKWMGPI 219  
DB 178 AYGAWINVHTEDGSEHSFFGSDRFHLIADLLOP--QPLP 216

RESULT 4

YIS7\_CABEL STANDARD; PRT; 225 AA.  
AC Q18973;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 26.0 kDa protein D2024.7 in chromosome IV.  
GN D2024.7  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderiinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Du 2.2; Gattung S.;  
RA Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RL -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U41011; AAA82289.1;  
DR WormPep; D2024.7; CE04296.  
DR InterPro: IPR004287; HCCA\_isomerase.  
DR Pfam; PF03046; HCCA\_isomerase; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 225 AA; 25951 MW; 23D0A3D6762B7232 CRC64;

Query Match 25.2%; Score 298.5; DB 1; Length 225;  
Best Local Similarity 33.0%; Pred. No. 8.4e-19;  
Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;

QY 3 PLPRTVELFVDVLSWLGFEILCRYQNIWNINLQRLPSLTITGIMKDSGNKPPGL--L 59  
DB 2 PNRRVVRFFDVISPYSGEGEITRRSVKWTPTQMKPFFAGVVRHTEN--PGLPLRI 59  
QY 60 PRKGLYANDLKLRLHLLQIPHFPKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
DB 60 PIKEYMHKDLFLSAQYWGIPRLPKOVNTNMLNTSSIVPQRILVASQSLDNVLMEDVAR 119  
QY 120 ELWMRVWSRNEDI-TEPQSILAAEKAGMSAEQAQGLLEKIATPKVKNLKETTEAACRY 178

```

Db 63 LKVDLQWQAQYGLVPLFPANYSRRNIGFYSGAQAAYVNV-----VF 110
QY 123 MRVMSRNDITEPOSILAA--AEKAGMSAEQAQGLLEKATPKVKNQKLTETEAACRYGA 180
Db 111 NAVM--CEGIAPDLESPLVSEKLGWDRSAFEHFLUSSNAATE---RYDEQTHAAIERKV 165
QY 181 FGLPITVAHVDGQTHMLFGSDRMELLALLG 211
Db 166 FGVP-TMFLGD---EMWGNDRFLMESAMG 192

RESULT 6
PTRA_ECOLI
AC P05458; P78106; STANDARD; PRT; 962 AA.
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Protease III precursor (EC 3.4.24.35) (Pitrilysin) (Protease pi).
OS PTR A OR PTR B2821.
OS Escherichia coli.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=87040734; PubMed=3534791;
RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emmerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli ptr gene
RL encoding protease III."
RN Nucleic Acids Res. 14:7695-7703(1986).
RN [2]
RN SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RN SEQUENCE OF 853-962 FROM N.A.
RX MEDLINE=87066729; PubMed=3537960;
RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RA Emmerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli recB gene."
RN Nucleic Acids Res. 14:8573-8582(1986).
RN [4]
RN SEQUENCE OF 1-296 FROM N.A.
RX STRAIN=K12;
RX MEDLINE=88005781; PubMed=3308636;
RA Claverie-Martin F., Diaz-Torres M.R., Kushner S.R.;
RT "Analysis of the regulatory region of the protease III (ptr) gene of
RL Escherichia coli K-12."
RN Gene 54:185-195(1987).
RN [5]
RN MUTAGENESIS, AND ACTIVE SITE.
RX MEDLINE=92237263; PubMed=1570301;
RA Becker A.B., Roth R.A.;
RT "An unusual active site identified in a family of zinc
RL metalloendopeptidases."
RN Proc. Natl. Acad. Sci. U.S.A. 89:3835-3839(1992).
RN [6]
RN MUTAGENESIS, AND ACTIVE SITE.
RX MEDLINE=93277449; PubMed=8099278;
RA Becker A.B., Roth R.A.;
RT "Identification of glutamate-169 as the third zinc-binding residue in
RL proteinase III, a member of the family of insulin-degrading
RN enzymes."
RL Biochem. J. 292:137-142(1993).
CC -!- FUNCTION: ENDOPEPTIDASE THAT DEGRADES SMALL PEPTIDES OF LESS THAN
CC 7 kDa, SUCH AS GLUCAGON AND INSULIN.

```

```

CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF 16-TYR-|-LEU-17 AND
CC 25-PHE-|-TYR-26 BONDS OF OXIDIZED INSULIN B CHAIN. ALSO ACTS ON
CC OTHER SUBSTRATES OF MW LESS THAN 7 kDa SUCH AS INSULIN AND
CC GLUCAGON.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY. BINDS ZINC.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16; ALSO KNOWN AS THE
CC INSULINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. Usage by and for commercial
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; X04581; CAA28249.1; -
CC EMBL; U29581; AAB40468.1; -
CC EMBL; AE000365; AAC75860.1; -
CC EMBL; X06227; CAA29576.1; -
CC EMBL; M17095; AAA24436.1; -
CC PIR; A29093; SNECPI.
CC MEROPS; M16.001; -
CC EC02DBASE; G095.0; 6TH EDITION.
CC Ecogene; EGI0786; ptrA.
CC InterPro; IPR001431; Peptidase_M16.
CC Pfam; PF00675; Peptidase_M16; 1.
CC PROSITE; PS00143; INSULINASE; 1.
CC Hydrolase; Metalloprotease; Magnesium; Zinc; Signal;
KW Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 962 PROTEASE III.
FT METAL 88 88 ZINC.
FT ACT_SITE 91 91
FT METAL 92 92 ZINC.
FT METAL 169 169 ZINC.
FT MUTAGEN 88 88 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.
FT MUTAGEN 91 91 E->Q: LOSS OF ACTIVITY.
FT MUTAGEN 92 92 H->R: LOSS OF ACTIVITY AND OF ZN-BINDING.
FT MUTAGEN 162 162 E->Q: 20% LOSS OF ACTIVITY.
FT MUTAGEN 169 169 E->Q: LOSS OF ACTIVITY AND OF ZN-BINDING.
FT MUTAGEN 204 204 E->Q: NO LOSS OF ACTIVITY.
FT CONFLICT 277 284 IIIHYVPA -> HYHSLRPW (IN REF. 4).
SQ SEQUENCE 962 AA; 107708 MW; 0558C68C2F1A0540 CRC64;

Query Match 7.6%; Score 89.5; DB 1; Length 962;
Best Local Similarity 26.3%; Pred. No. 5.4;
Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;

QY 126 WSRNDEITEPOSILAAAEKAGMSAEQA-----QGLEKIATPKVKNQ 167
Db 731 WCRNKDVVVKKQSVTFEKAGNSTDSALAAVFVPTGYDEYTSAYSSLLGQIVQPFYNO 790
QY 168 LKETTEAACRYGAFGLPITVAHVDGQTHMLFGSDRMELLALLGKRWGPIPPA 221
Db 791 LR--TEEOGLYAVFAFFMSVGRWGWFLLQSDNQ---PSFLWERYKAFFPTA 839

RESULT 7
NAHD_PSESP
ID NAHD_PSESP STANDARD; PRT; 212 AA.
AC Q52462;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2-hydroxychromene-2-carboxylate isomerase (HCCA isomerase).
GN DOXJ
OS Pseudomonas sp. (strain C18).
OG Plasmid.
CC Bacteria; Proteobacteria.

```

STRAIN=CV. COLUMBIA;  
Mourad G., Emerick R., Marion A., Smith A.;  
"Cloning and sequencing of a cDNA encoding threonine dehydratase/  
deaminase of Arabidopsis thaliana."; (In) Plant Gene Register PGR98-199.  
[2]  
SEQUENCE FROM N.A.  
Mourad G.S., Smith A.M.;  
"Molecular characterization of the genomic clone, including the  
promoter sequences, of threonine dehydratase/deaminase from  
Arabidopsis thaliana."; to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A.  
STRAIN=CV. Gmlib;  
MEDLINE=20144028; PubMed=10677454;  
Mourad G., Emerick R., Smith A.;  
"Molecular cloning and sequencing of a cDNA encoding an isoleucine  
feedback insensitive threonine dehydratase/deaminase of Arabidopsis  
thaliana line Gmlib."; (In) Plant Gene Register PGR00-020.  
[4]  
SEQUENCE FROM N.A.  
STRAIN=CV. COLUMBIA;  
MEDLINE=21016720; PubMed=11130713;  
RANANOUAT M., LEMCKE K., RIEGER M., ANSORGE W., UNSELD M.,  
SALANOUAT M., BLOECKER H., PEREZ-ALONSO M., OBERMALER B.,  
FARTMANN B., VALLE G., BLOECKER H., GRIVELL L.A., MACHE R., PUIGDOMENEC P.,  
DELSENY M., BOUTRY M., CHOISNE N., ARTIGUENAVE F., ROBERT C., BROTTIER P.,  
DE SIMONE V., CATOLICO L., WEISSENBACH J., SAURIN W., QUETIER F.,  
WINCKER P., CATOLICO L., WEISSENBACH J., SAURIN W., QUETIER F.,  
SCHAEFER M., MUELLER-AUER S., GABEL C., FUCHS M., RENES V.,  
WURMBACH E., DRZONEK H., ERFLE H., JORDAN N., BANGERT S.,  
WIEDELMANN R., KRANZ H., VOSS H., HOLLAND R., BRANDT P., NYAKATURA G.,  
VEZZI A., D'ANGELO M., PALLAVICINI A., TOPPO S., SIMONATI B.,  
CONRAD A., HORNISCHER K., KAUER G., LOEHNER T.-H., NORDSTIEK G.,  
REICHELDT J., SCHAFER M., SCHOEN G., BARGUES M., TEROL J., CLIMENT J.,  
NAVARRO P., COLLADO C., PEREZ-PEREZ A., OTTENWAELEDER B., DUCHEMIN D.,  
COOKE R., LAUDIE M., BERGER-LLAURO C., PURNELLE B., MASUY D.,  
DE HAAN M., MAARSE A.C., ALCARAZ J.-P., COTTET A., CASACUBERTA E.,  
MONFORT A., ARGILLOU A., FLORES M., LIGUORI R., VITALE D.,  
MANNHAUPT G., HAASE D., SCHOOF H., RUDD S., ZACCARIA P., MEWES H.-W.,  
MAYER K.F.X., KAUL S., TOWN C.D., KOO H.L., TAILLON L.J., JENKINS J.,  
RIZZO M., WALT S., UTTERBACK T., FUJII C.F., SHEA T.P.,  
CREASY T.H., HAAS B., MATTI R., WU D., PETERSON J., VAN AKEN S.,  
PAL G., MILITSCHER J., SALLERS P., GILL J.E., FEIDBIYUM T.V.,  
PREUSS D., LIN X., NIEMAN W.C., SALZBERG S.L., WHITE O., VENTER J.C.,  
FRASER C.M., KANEKO T., NAKAMURA Y., SATO S., KATO T., ASAMIZU E.,  
SASAMOTO S., KIMURA T., IDESAWA K., KAWASHIMA K., KISHIDA Y.,  
KIKYOKAWA C., KOHARA M., MATSUMOTO M., MATSUNO A., MURAKI A.,  
NAKAYAMA S., NAKAZAKI N., SHINPO S., TAKEUCHI C., WADA T.,  
WATANABE A., YAMADA M., YASUDA M., TABATA S.;  
"Sequence and analysis of chromosome 3 of the plant Arabidopsis  
thaliana."; Nature 408:820-822(2000).  
-1- FUNCTION: Catalyzes the formation of alpha-ketobutyrate from  
threonine in a two-step reaction. The first step is a dehydration  
of threonine, followed by rehydration and liberation of ammonia.  
-1- CATALYTIC ACTIVITY: L-threonine + H<sub>2</sub>O = 2-oxobutanate + NH<sub>3</sub> +  
H<sub>2</sub>O.  
-1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
-1- ENZYME REGULATION: Allosterically inhibited by isoleucine. Strain  
Gmlib is isoleucine feedback insensitive and is resistant to the  
antimetabolite L-0-methylthreonine.  
-1- PATHWAY: Isoleucine biosynthesis; first step.  
-1- SUBCELLULAR LOCATION: Chloroplast (By similarity).  
-1- SIMILARITY: BELONGS TO THE SERINE/THREONINE DEHYDRATASE FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial/  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)



CC CRNA containing 6-isopentenyladenosine.  
 CC -1- SIMILARITY: BELONGS TO THE IPP TRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001238; AAC65611.1; -  
 CC TIGR: TP06377; 002627; IPPT.  
 CC InterPro: IPR002627; IPPT.  
 CC Pfam: PF01715; IPPT; 1.  
 CC ProDom: PD004674; IPPT; 1.  
 CC Transfaser: Nucleotidyltransferase; tRNA processing; ATP-binding;  
 CC Complete proteome.  
 CC NP\_BIND 27 34 ATP (POTENTIAL).  
 CC SEQUENCE 316 AA; 36174 MW; E4BE144EE6ED5208 CRC64;  
 SQ  
 Query Match 7.1%; Score 84.5; DB 1; Length 316;  
 Best Local Similarity 23.3%; Pred. No. 3.8; Mismatches 45; Gaps 8;  
 Matches 41; Conservative 30;  
 QY 13 DVLSVSWLGFELCRQYQNIWNLQRLPSLITGIMKDSGNKPGGLPRKGLYMANDLKL 72  
 DB 80 DCDPVE--EYNVFRFOQAVYGI-----VPSIL-----RAHKVPIIVGTGLYLD---AV 124  
 QY 73 LRHLQIPIHPKDFLSVLMKSGLSAM-----RELTAVALN- 108  
 DB 125 LRQYALVPE--RNQALRASLRGASLSHMRVAVYFSKDSHAVHNKTDLEDPARLMRAIEIA 183  
 QY 109 -----EHEPMLEKASRELWMYRSNEDITEPOSILAAAEKAGMSAEQAQGLLEKIA 160  
 DB 184 VFHATHPELLOQA--RETPRPMRAKYVGYQYPRSMRLPARIRARLEQRIGRGLIEVA 238  
 RESULT 11  
 S3BL\_YEAST  
 ID S3BL\_YEAST STANDARD; PRT: 971 AA.  
 AC P49955;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE U2 snRNP component HSH155.  
 GN HSH155 OR YMR288W OR YMR021.14.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: contacts pre-mRNA on both sides of the branch site early  
 CC in spliceosome assembly (By similarity).  
 CC -1- SUBUNIT: Subunit of the U2 snRNP (By similarity).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SF3B1 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 9 HEAT REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z49704; CAA89786.1; -  
 DR

DR SGD: S0004901; HSH155.  
 DR InterPro: IPR000357; HEAT\_REPEAT.  
 DR PROSITE: PS50077; HEAT\_REPEAT; mRNA\_splicing; Nuclear protein; Repeat.  
 KW Spliceosome; mRNA processing; HEAT 1.  
 FT REPEAT 199 237  
 FT REPEAT 273 310 HEAT 2.  
 FT REPEAT 350 387 HEAT 3.  
 FT REPEAT 513 550 HEAT 4.  
 FT REPEAT 596 633 HEAT 5.  
 FT REPEAT 680 717 HEAT 6.  
 FT REPEAT 722 759 HEAT 7.  
 FT REPEAT 792 829 HEAT 8.  
 FT REPEAT 832 870 HEAT 9.  
 SQ SEQUENCE 971 AA; 110027 MW; 27D26EA4252A788E2 CRC64;  
 Query Match 7.1%; Score 84.5; DB 1; Length 971;  
 Best Local Similarity 21.7%; Pred. No. 15;  
 Matches 36; Conservative 31; Mismatches 64; Indels 35; Gaps 7;  
 QY 5 PRTVEFVYDLSVSWLGFELCRQYQNIWNLQRLPSLITGIMKDSGNKPGGLPR-KG 63  
 DB 387 PYGTEVFNVLPE-----LWKGIRSHRGKVLSEFLKAVGMIPLMDPEYAG 432  
 QY 64 LYMANDLKLRRHQLQIHPKDFLSVLMKSGLSAMRFLTAVNLEHPEML-EKASRELW 122  
 DB 433 YTTTEAMRIIRREEDSPDDEMKKTIILLVLOK---CSAVESIT-----PKFLREETAEPEFF 484  
 QY 123 MRVSRNEDITEP-----OSILAAAEKAGMSAEQAQGLLEKIATP 162  
 DB 485 QKFWRRVALDRLPKNVVYTYTTLAKKLGCSY-----TIDKLLTP 525  
 RESULT 12  
 CATA\_LISSE  
 ID CATA\_LISSE STANDARD; PRT: 488 AA.  
 AC P24168;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Catalase (EC 1.11.1.6).  
 GN KAT.  
 OS Listeria seeligeri.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Listeria.  
 OC NCBI\_TaxID=1640;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91317736; PubMed=1860824;  
 RA Haas A., Brehm K., Krefl J., Goebel W.;  
 RT "Cloning, characterization, and expression in Escherichia coli of a  
 RT gene encoding Listeria seeligeri catalase, a bacterial enzyme highly  
 RT homologous to mammalian catalases";  
 RL J. Bacteriol. 173:5159-5167(1991).  
 CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES  
 CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.  
 CC -1- COFACTOR: HEME GROUP.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M75944; AAB53655.1; -  
 DR PIR: A40367; A40367.  
 DR HSP: P21179; 1CF9.  
 DR InterPro: IPR002226; Catalase.  
 DR

DR Pfam: PF00199; catalase; 1.  
 DR PRINTS: PR00067; CATALASE.  
 DR ProDom: PD000510; Catalase; 1.  
 DR PROSITE: PS00437; CATALASE; 1.  
 DR PROSITE: PS00438; CATALASE; 2.  
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.  
 FT ACT\_SITE 55  
 FT BINDING 128 338  
 FT BINDING 338 338  
 SQ SEQUENCE 488 AA; 55869 MW; 50A3251469681EBB CRC64;  
 Query Match 7.0%; Score 83.5; DB 1; Length 488;  
 Best Local Similarity 21.3%; Pred. No. 7.9;  
 Matches 52; Conservative 37; Mismatches 76; Indels 79; Gaps 12;  
 QY 9 ELFDVLPSP-----YSLGFEILCYQNIWNINLQRPRLITIMKDSGNK 54  
 Db DLYVQVLPKLDNDYFNPLDKDW--FEDVFPYHVGVNMTLARNPDNIFAETESVGFN 315  
 QY 55 P-----PGLLPKRG-----LYMANDLKLRL--HHLQIPHFPPKDFLSVMLEKGSLSAMR 101  
 Db DLYVQVLPKLDNDYFNPLDKDW--FEDVFPYHVGVNMTLARNPDNIFAETESVGFN 315  
 QY 316 PGVLVPGMLPSDRVLQGRLEYSYDQTHRVGPNYLPINSPKTPVDNNDGDMPPKQ 375  
 Db DLYVQVLPKLDNDYFNPLDKDW--FEDVFPYHVGVNMTLARNPDNIFAETESVGFN 315  
 QY 102 FLTAVNLF-----HPE-----MLEKA-----SRELWVRWSRNE 130  
 Db DLYVQVLPKLDNDYFNPLDKDW--FEDVFPYHVGVNMTLARNPDNIFAETESVGFN 315  
 QY 376 QTSSINYPNSVDTEPKENPAVIEPEQIEIRGDISGRVLAEPKNNFNGHAKVWKR----- 429  
 Db DLYVQVLPKLDNDYFNPLDKDW--FEDVFPYHVGVNMTLARNPDNIFAETESVGFN 315  
 QY 131 DITEPQSTLAAAEKAGM---SAEQAGLLEKIATPKVKNOLKETTEAACRGVGFGLPTIV 187  
 Db DLYVQVLPKLDNDYFNPLDKDW--FEDVFPYHVGVNMTLARNPDNIFAETESVGFN 315  
 QY 430 -----YSDAERAAALVKNIVDDWEGVREDIKIRLNRFYQVEPEFAERVA--GTGINL 480  
 Db DLYVQVLPKLDNDYFNPLDKDW--FEDVFPYHVGVNMTLARNPDNIFAETESVGFN 315  
 QY 188 A-HV 190  
 Db 481 AEHV 484

RESULT 13  
 NCD\_DROME  
 ID NCD\_DROME STANDARD; PRT; 700 AA.  
 AC P20480; Q9VAG8;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Claret segregational protein.  
 GN NCD OR CA(ND) OR CG7831.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN - [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R, AND CANTON-S; TISSUE=Ovary;  
 RX MEDLINE=90231469; PubMed=1691829;  
 RA Endow S.A., Henikoff S., Soler-Niedziela L.;  
 RT "Mediation of meiotic and early mitotic chromosome segregation in  
 RL Drosophila by a protein related to kinesin."  
 RL Nature 345:81-83(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champ E., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman D.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouch J., Brockstein P., Brothier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de  
 RA de Fablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE OF 16-700 FROM N.A.  
 RX MEDLINE=90275618; PubMed=2140958;  
 RA McDonald H.B., Goldstein L.S.B.;  
 RT "Identification and characterization of a gene encoding a  
 RT kinesin-like protein in Drosophila."  
 RL Cell 61:991-1000(1990).  
 RN [4]  
 RP MOTOR DIRECTIONALITY.  
 RX MEDLINE=91043032; PubMed=2146510;  
 RA Walker R.A., Salmon E.D., Endow S.A.;  
 RT "The Drosophila claret segregation protein is a minus-end directed  
 RT motor molecule."  
 RL Nature 347:780-782(1990).  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94155838; PubMed=8112290;  
 RA Lockhart A., Cross R.A.;  
 RT "Origins of reversed directionality in the ncd molecular motor."  
 RL EMBO J. 13:751-757(1994).  
 RN [6]  
 RP MUTANT ALLELE NCD(D).  
 RX MEDLINE=91122049; PubMed=1825056;  
 RA Komma D.J., Horne A.S., Endow S.A.;  
 RT "Separation of meiotic and mitotic effects of claret  
 RT non-disjunctional on chromosome segregation in Drosophila."  
 RL EMBO J. 10:419-424(1991).  
 RN [7]  
 RP CHARACTERIZATION OF MUTANT ALLELE NCD(D).  
 RX MEDLINE=96283629; PubMed=8670831;  
 RA Moore J.D., Song H., Endow S.A.;  
 RT "A point mutation in the microtubule binding region of the ncd motor  
 RT protein reduces motor velocity."  
 RL EMBO J. 15:3306-3314(1996).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 335-700.  
 RX MEDLINE=96195067; PubMed=8606780;  
 RA Sablin E.P., Kull F.J., Cooke R., Vale R.D., Fletterick R.J.;  
 RT "Crystal structure of the motor domain of the kinesin-related motor  
 RT ncd."  
 RL Nature 380:555-559(1996).  
 CC - FUNCTION: NCD IS REQUIRED FOR NORMAL CHROMOSOMAL SEGREGATION IN  
 CC MEIOSIS, IN FEMALES, AND IN EARLY MITOTIC DIVISIONS OF THE EMBRYO.  
 CC "THE NCD MOTOR ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S MINUS  
 CC END."



Tue May 21 08:04:02 2002

```

CC CC MEDLINE-8909327; PubMed-32111750;
CC RA Gerdes H.-H., Philipps E., Huttner W.B.;
CC RT "The primary structure of rat secretogranin II deduced from a cDNA
CC sequence.";
CC RL Nucleic Acids Res. 16:11811-11811(1988).
CC [2]
CC RP SEQUENCE FROM N.A.
CC MEDLINE-93309708; PubMed-8321414;
CC RA Kakar S.S., Wei N., Mulchahey J.J., Leboeuf R.D., Neill J.D.;
CC RT "Regulation of expression of secretogranin II mRNA in female rat
CC pituitary and hypothalamus.";
CC RL Neuroendocrinology 57:422-431(1993).
CC CC -!- FUNCTION: SECRETOGHRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE
CC PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.
CC CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY
CC GRANULES.
CC CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
CC CC -!- SIMILARITY: BELONGS TO THE CHROMOGHRANIN / SECRETOGHRANIN PROTEIN
CC FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC EMBL; X13618; CAA31950.1; -
CC DR EMBL; M93669; AAM42135.1; -
CC DR PIR; S02180; S02180.
CC DR InterPro; IPR001990; Granin.
CC DR Pfam; PF01271; Granin; 1.
CC DR PROSITE; PS00422; GRANINS_1; 1.
CC KW Sulfation; Cleavage on pair of basic residues: Calcium-binding;
CC Signal.
CC FT SIGNAL. 1 30 SECRETOGHRANIN II.
CC FT CHAIN 31 619 UNKNOWN ACTIVITY PEPTIDE (PROBABLE).
CC FT PEPTIDE 184 216 SUFFATION (BY SIMILARITY).
CC FT MOD_RES 153 153
CC FT SEQUENCE 619 AA; 71031 MW; 27CB75B4F25A38D1 CRC64;
CC SQ
Query Match 6.9%; Score 81.5; DB 1; Length 619;
Best Local Similarity 24.8%; Pred. No. 16;
Matches 41; Conservative 16; Mismatches 57; Indels 51; Gaps 8;
QY 13 DVLSPSYSLGFEILCRYQNIWNINQLRPSLTIGIMKSGNKKPPGILLPRKGLYMANDLKL 72
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 105 DVLSDEWM-----RIIEALRQAEENPPPSALKENKPYALNLEK- 143
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 LRHLQIPHPKDFLSVLMLEKSGLSAMRFLTAVNLEHPMLEKASRELMMRVMSRVEDI 132
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 144 -----NPFVDPDDYETQOWPERKLEKMRP-----PLMYEENSRE----NPFKRTNEI 187
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 TE-----QSILAAAEKAGMSAQOGLLEKIATPKVKNQKETTE 173
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 188 VEEQYTPQS-LATLESVFQE-----LGKLTGP--SNQKRERVD 222
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 15
DPOL_ADE12 STANDARD; PRT; 1061 AA.
ID DPOL_ADE12
AC P06538;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Secretogranin II precursor (SGII) (Chromogranin C).
GN SG2 OR CHGC OR SCG-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

```

CC CC MEDLINE-8909327; PubMed-32111750;
CC RA Gerdes H.-H., Philipps E., Huttner W.B.;
CC RT "The primary structure of rat secretogranin II deduced from a cDNA
CC sequence.";
CC RL Nucleic Acids Res. 16:11811-11811(1988).
CC [2]
CC RP SEQUENCE FROM N.A.
CC MEDLINE-93309708; PubMed-8321414;
CC RA Kakar S.S., Wei N., Mulchahey J.J., Leboeuf R.D., Neill J.D.;
CC RT "Regulation of expression of secretogranin II mRNA in female rat
CC pituitary and hypothalamus.";
CC RL Neuroendocrinology 57:422-431(1993).
CC CC -!- FUNCTION: SECRETOGHRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE
CC PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.
CC CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY
CC GRANULES.
CC CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.
CC CC -!- SIMILARITY: BELONGS TO THE CHROMOGHRANIN / SECRETOGHRANIN PROTEIN
CC FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC EMBL; X52814; CAA36998.1; -
CC DR EMBL; M33932; AAM28716.1; -
CC DR EMBL; AE003771; AAF56942.1; -
CC DR EMBL; X57475; CAA40713.1; -
CC DR PIR; A35624; A35624.
CC DR PIR; S09748; S09748.
CC DR HSP; P17119; 3KAR.
CC DR FlyBase; FBgn0002924; ncd.
CC DR InterPro; IPR001752; kinesin.
CC DR Pfam; PF00225; kinesin; 1.
CC DR PRINTS; PR00380; KINESINHEAVY.
CC DR SMART; SM00129; KTSC; 1.
CC DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
CC KW Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
CC Meiosis; Mitosis.
CC KW DOMAIN 196 346 COILED COIL (POTENTIAL).
CC FT DOMAIN 347 700 KINESIN-MOTOR (BY SIMILARITY).
CC FT NP_BIND 434 441 ATP (BY SIMILARITY).
CC FT MOTAGEN 556 556 V->F: IN NCD(D); REDUCES MOTOR VELOCITY.
CC FT CONFLICT 697 697 S -> N (IN REF. 1).
CC FT SEQUENCE 700 AA; 77473 MW; ADE043CBCE7FD561 CRC64;
CC SQ
Query Match 7.0%; Score 83.5; DB 1; Length 700;
Best Local Similarity 22.4%; Pred. No. 12;
Matches 52; Conservative 37; Mismatches 70; Indels 73; Gaps 12;
QY 1 MGPIPTVELFDVLSPSYSLG--FEILCRYQNIWNINQLRPSLTIGIMKD-----SG 52
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 451 VGVIPRTVDLFDSDRGVNLGWELKATFLEIYN---EVLYDLLSNEQKMEIRMAKN 507
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 53 NKPPCLLPKGLYMAN---DLKLLRHHLQIPHPKDFLSVLMLEKSGLSAMRFLTAVNLE 109
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 508 NK-----NDIYVSNITEETVLDPNHLRLHMTAK-----MNRATAS--TAGN-- 547
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 HPMELEKASRELMMRVMSRVEDITEQSILAAAEKAGMS-----AEQAQGLLEK 158.
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 548 -----ERS-----SRSHAVTKLEIGRHAQKQISVGSINLVLAGSESPKSTRM 593
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 IATPKVKNQKETTEAACRYGAFGLPTTVAHVQDQTHLMFGSDRMELLALL 210
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 594 TETKNINRSLSELTN-----VILALLQODHIPY---RNSKLTHLL 631
||| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 14
SG2_RAT STANDARD; PRT; 619 AA.
ID SG2_RAT
AC P10362;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Secretogranin II precursor (SGII) (Chromogranin C).
GN SG2 OR CHGC OR SCG-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

```
RP SEQUENCE FROM N.A.
RX MEDLINE-94076430; PubMed-8254750;
RA Sprengel J., Schmitz B., Heuss-Neitzel D., Zock C., Doerfler W.;
RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative
RL functional analysis";
RN J. Virol. 68:379-389(1994).
RP SEQUENCE FROM N.A.
RX MEDLINE-87106854; PubMed-3803925;
RA Shu L., Hong J.S., Wei Y.-F., Engler J.A.;
RT "Nucleotide sequence of the genes encoded in early region 2b of human
RL adenovirus type 12";
CC Gene 46:187-195(1986).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + {DNA}(N).
CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73487; CAA51882.1; -
DR EMBL; M14785; AAA2478.1; ALT_INIT.
DR PIR; A25770; DJAD12.
DR PIR; S33933; S33933.
DR InterPro; IPR002064; DNA_pol_B.
DR Pfam; PF00136; DNA_pol_B; 1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding.
FT CONFLICT 32 32 R -> S (IN REF. 2).
FT CONFLICT 162 162 V -> L (IN REF. 2).
FT CONFLICT 181 182 LQ -> VN (IN REF. 2).
FT CONFLICT 461 461 S -> T (IN REF. 2).
FT CONFLICT 575 575 L -> F (IN REF. 2).
FT CONFLICT 892 892 S -> T (IN REF. 2).
FT CONFLICT 1030 1030 K -> M (IN REF. 2).
SQ SEQUENCE 1061 AA; 121727 MW; 33FBA89C33065C08 CRC64;

Query Match      6.9%; Score 81.5; DB 1; Length 1061;
Best Local Similarity 21.8%; Pred No. 30;
Matches 50; Conservative 35; Mismatches 77; Indels 67; Gaps 12;
Qy 1 MGPLPRTVELF--YDVLSPYSWLGFEILCRYQNIWNINLQRPISLITGIMKDSGNKPPGL 58
Db 133 IGSHPRTERLFTYDV-ETYTWG-----AFGKQLVPEML--VMKLSG----- 172
Qy 59 LPRKGLYMANDLKLRLHQLIPI-----HPKQ---FLSVMLEKGSLSAMRFLTAVNLEH 110
Db 173 -----DDNLVKHALQLALELQWDQWEKSTTFYCLTPEKMKV--GQGFRTYRN--- 218
Qy 111 PEMLEKASRELMRVNSREDITEPOSILAAAEKAGMSAQOGLLEKIATPKVKNOLKE 170
Db 219 -RLQTSLATDLMMTFLQKNPHLSQ-----WAQENGIVALEDLSEYDLKRAPAIKGEPR- 271
Qy 171 TTEACRYGAFGLPITVAHVGDGTHMLFGSDRMELLALLGKWMGPIP 219
Db 272 -----FVELYIVG--HNINGFDEIVLAAQVINNRDLDPGP 304
```

Search completed: May 20, 2002, 08:42:28  
Job time: 195 sec

Tue May 21 08:04:00 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:33:28 ; Search time 30.57 Seconds  
(without alignments)  
821.154 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 1185  
Sequence: 1 MGPLPRTVELFDVLSYSW.....AHLGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapox 10.0 , Gapext 0.5  
Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match %	Length	ID	Description
1	1185	100.0	226	21	Novel human glutathione S-transferase, human; GSTS; cancer; immune disorder; gene therapy; diagnosis; treatment; drug screening.
2	1185	100.0	256	20	Novel human glutathione S-transferase, human; GSTS; cancer; immune disorder; gene therapy; diagnosis; treatment; drug screening.
3	820.5	69.2	300	21	Human endometrium
4	700	59.1	131	21	Human OREF3062
5	496	41.9	97	21	Human secreted pro
6	387	32.7	97	21	Human secreted pro
7	326.5	27.6	107	22	Gene 45 human secr
8	89.5	7.6	1086	22	Human reproductive
9	89	7.5	545	20	Novel human diago
10	89	7.5	545	20	Mutant threonine d
11	89	7.5	590	20	Feedback insensiti

12	89	7.5	592	20	AAAY32939	Mutant threonine d
13	89	7.5	592	20	AAAY32948	Mutant threonine d
14	89	7.5	592	20	AAAY32951	Wild type threonin
15	89	7.5	592	20	AAAY05702	Arabidopsis wild-t
16	89	7.5	592	20	AAAY05703	Feedback insensiti
17	89	7.5	600	20	AAAY32952	Feedback insensiti
18	89	7.5	609	20	AAAY32940	Mutant threonine d
19	89	7.5	609	20	AAAY05704	Feedback insensiti
20	88.5	7.5	395	22	AAG90974	C glutamicum prote
21	88.5	7.5	399	22	AAG92476	C glutamicum prote
22	83.5	7.0	308	17	AAW04266	Inosine-guanosine
23	83.5	7.0	539	20	AAAY32942	Mutant threonine d
24	83.5	7.0	539	20	AAAY05706	Feedback insensiti
25	83.5	7.0	700	22	ABW58943	Drosophila melanog
26	82.5	7.0	532	22	ABG05274	Novel human diago
27	81	6.8	586	21	AAAB1686	Bacteriophage Dp-1
28	81	6.8	911	20	AAAY5957	Mouse STP20-relate
29	80.5	6.8	339	20	AAAY33825	Amino acid sequenc
30	80.5	6.8	532	20	AAAY05707	Mutant threonine d
31	80.5	6.8	662	22	ABG17704	Feedback insensiti
32	80.5	6.8	96	21	AAAB29476	Novel human diago
33	80	6.8	245	22	AAAG92570	Burkholderia cis 3
34	80	6.8	245	22	AAAB76520	C glutamicum prote
35	80	6.8	704	22	ABW60264	Corynebacterium gl
36	80	6.8	1073	18	AAW32063	Drosophila melanog
37	79.5	6.7	1073	19	AAW37371	Human ST receptor
38	79.5	6.7	279	21	AAAG43853	Arabidopsis thalia
39	79	6.6	584	22	ABG30321	Novel human diago
40	78.5	6.6	584	22	ABW63453	Drosophila melanog
41	78.5	6.6	633	22	ABW63453	Amino acid sequenc
42	78	6.6	341	20	AAAY37006	Novel human diago
43	77.5	6.5	887	22	ABG03067	Novel human secret
44	77	6.5	172	22	AAU30961	Novel human diago
45	77	6.5	385	22	ABG18840	Novel human diago

ALIGNMENTS

RESULT 1  
AAAY77499  
ID AAAY77499 standard; Protein; 236 AA.  
XX  
AC AAAY77499;  
XX  
DT 05-JUN-2000 (first entry)  
XX  
DE Novel human glutathione S-transferase, GSTS.  
XX  
KW Glutathione S-transferase; human; GSTS; cancer; immune disorder;  
KW gene therapy; diagnosis; treatment; drug screening.  
XX  
OS Homo sapiens.  
XX  
PN US6030809-A.  
XX  
PD 29-FEB-2000.  
XX  
PF 25-NOV-1997; 97US-0978174.  
XX  
PR 25-NOV-1997; 97US-0978174.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Hillman JL, Shah P, Lal P, Corley NC;  
XX  
DR WPI; 2000-205204/18.  
XX  
DR N-PSDB; AA025599.  
XX  
PT Isolated nucleic acid encoding glutathione S-transferase useful in the  
PT production of agents for preventing, diagnosing and treating diseases  
PT associated with cell proliferation -

PS Claim 8; Fig 1A-C; 27pp; English.

XX This sequence represents a novel human glutathione S-transferase, GSTS.  
CC Nucleotide sequences encoding GSTS were initially isolated from a  
CC urologic CDNA library, and subsequently extended using cDNA libraries  
CC derived from other tissues, such as brain or bladder. The present  
CC sequence is encoded by a consensus cDNA, GSTS, and nucleotides which  
CC encode it may be used in the prevention, treatment and diagnosis of  
CC diseases associated with inappropriate GSTS expression, such as cancers  
CC and immune disorders. Nucleotides which encode GSTS may be used in gene  
CC therapy to treat disorders associated with reduced expression or activity  
CC of GSTS, and in antisense therapy for disorders associated with increased  
CC GSTS expression or activity. They may also be used for the recombinant  
CC production of GSTS, and as a source of probes and primers to detect and  
CC quantitate the presence of similar nucleic acid sequences, particularly  
CC for the diagnosis of GSTS-associated disorders. GSTS proteins may be used  
CC as antigens in the production of antibodies against GSTS and in assays to  
CC identify modulators (agonists and antagonists) of GSTS expression and  
CC activity. The anti-GSTs antibodies and GSTS antagonists may also be used  
CC to downregulate GSTS expression and activity. Antagonists of GSTS  
CC expression and function may be used to treat immune disorders (e.g., AIDS,  
CC anaemia, asthma, Crohn's disease, irritable bowel syndrome, multiple  
CC sclerosis, osteoarthritis, microbial infections) and cancers (e.g.,  
CC leukaemia, lymphoma, melanoma, and cancers of the breast, prostate,  
CC liver, lung and brain). The anti-GSTs antibodies may also be used as  
CC diagnostic agents.  
XX Sequence 226 AA;

Query Match 100.0%; Score 1185; DB 21; Length 226;  
Best Local Similarity 100.0%; Pred. No. 1.3e-121;  
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
DB 1 mgplprtvelfdvlsyslwgfeilcryqniwnlnqlrpslittgimkdsngkppglip 60  
QY 61 RKGLYMANDKLLRHHLQIPIHFPKDFLSVLEKGSLSAMRFLTAVNLEHPMLEKASRE 120  
DB 61 rkglymandkllrhhllqpihfpkdfslvlekgslsamrfltavnl ehpmlekasre 120  
QY 121 LWMRVWSRNEDITEPOSILAAAEKAGMSAEQAGGILLEKIATPKVKNLKETTEAACRYGA 180  
DB 121 lwmrwvsnrediteposilaaekagmsaeqaggllekiatpkvknqketteaacryga 180  
QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLLGKWMGPPIPPAVNARL 226  
DB 181 fglpityvahvdgqthmlfgsdrmlahllgkwmgpippavnarl 226

RESULT 2  
AAI59988  
ID AAY59988 standard; Protein; 256 AA.  
XX AC AAY59988;  
XX 31-JAN-2000 (first entry)  
XX Human endometrium tumour EST encoded protein 48.  
XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
KW treatment; uterine; gene therapy; expressed sequence tag.  
XX Homo sapiens.  
XX DE19817948-A1.  
XX 21-OCT-1999.  
XX 17-APR-1998; 98DE-1017948.  
XX 17-APR-1998; 98DE-1017948.  
PR

XX

PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;  
PI WPI; 1999-591957/51.  
XX N-PSDB; AA241996.  
XX New nucleic acid sequences expressed in uterine cancer tissues, and  
XX derived polypeptides for treatment of uterine and endometrial cancer  
XX and identification of therapeutic agents -  
XX Claim 23; Page 294; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for  
CC generation of specific antibodies. (A) are identified by assembling ESTs  
CC (expressed sequence tags) from a particular tissue type before comparison  
CC of expression patterns. This allows a significantly longer fragment of  
CC the gene to be revealed, so should reduce the number of failures  
CC associated with the fact that ESTs from different libraries may represent  
CC different parts of the same unknown gene, distorting the estimated  
CC frequency of occurrence in a particular tissue. AAY59941-Y60328 represent  
CC protein fragments encoded by the human endometrium tumour cDNA library  
CC derived EST fragments represented in AA241981-242121.

Sequence 256 AA;

Query Match 100.0%; Score 1185; DB 20; Length 256;  
Best Local Similarity 100.0%; Pred. No. 1.6e-121;  
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
DB 31 mgplprtvelfdvlsyslwgfeilcryqniwnlnqlrpslittgimkdsngkppglip 90  
QY 61 RKGLYMANDKLLRHHLQIPIHFPKDFLSVLEKGSLSAMRFLTAVNLEHPMLEKASRE 120  
DB 91 rkglymandkllrhhllqpihfpkdfslvlekgslsamrfltavnl ehpmlekasre 150  
QY 121 LWMRVWSRNEDITEPOSILAAAEKAGMSAEQAGGILLEKIATPKVKNLKETTEAACRYGA 180  
DB 151 lwmrwvsnrediteposilaaekagmsaeqaggllekiatpkvknqketteaacryga 210  
QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLLGKWMGPPIPPAVNARL 226  
DB 211 fglpityvahvdgqthmlfgsdrmlahllgkwmgpippavnarl 256

RESULT 3  
AAB43298  
ID AAB43298 standard; Protein; 300 AA.  
XX AC AAB43298;  
XX 08-FEB-2001 (first entry)  
XX Human ORF3062 polypeptide sequence SEQ ID NO:6124.

XX Human; open reading frame; ORF3; detection; cytostatic; hepatotropic;  
KW vulnery; antipsoriatic; antiparkinsonian; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
KW antiviral; antibacterial; antifungal; antineoplastic; antithyroid;  
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertenston;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;



Db 121 lmrwvwrned 131

RESULT 5  
 AAB34972  
 ID AAB34972 standard; Protein; 97 AA.  
 XX  
 AC AAB34972;  
 XX  
 DT 26-JAN-2001 (first entry)  
 XX  
 DE Human secreted protein sequence encoded by gene 45 SEQ ID NO:176.  
 XX  
 KW Human; secreted protein; neuroprotective; cytostatic; cardioactive;  
 KW immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic;  
 KW antinefactive; gynaecological; antibacterial; neural disorder; cancer;  
 KW immune disease; reproductive disorder; proliferative disease;  
 KW gastrointestinal disease; wound healing; infectious disease;  
 KW food additive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200056766-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 16-MAR-2000; 2000WO-US06824.  
 XX  
 PR 19-MAR-1999; 99US-0125359.  
 XX  
 PD 03-DEC-1999; 99US-0168664.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 DR WPI: 2000-594574/56.  
 DR N-PSDB; AAC60069.  
 XX  
 PT Human secreted proteins and gene sequences encoding them, useful for  
 PT detection, prevention, and treatment of various disorders such as  
 PT cancer and immune system disorders -  
 XX  
 PS Disclosure; Page 83; 442pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC60025-C60071 encode the human  
 CC secreted proteins represented in AAB34854-B34900. Sequences  
 CC AAB34901-B34976 are fragments of proteins encoded by the genes, and also  
 CC proteins with which they share sequence homology. The proteins have  
 CC activities based on the tissues in which their encoding genes are  
 CC expressed. Examples of the proteins activities include: neuroprotective;  
 CC - cytostatic; cardioactive; immunomodulatory; general muscular activity;  
 CC antinefactive; gynaecological; and antibacterial. The human secreted  
 CC proteins, polynucleotides, antagonists and antagonists of the invention  
 CC may be useful in treating, preventing and/or diagnosing various  
 CC diseases, disorders and conditions such as neural, immune, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal and  
 CC proliferative disorders and cancer. They may also be used in the  
 CC treatment of wounds, and infectious diseases. The polypeptides may be  
 CC used as a food additive or preservative to increase storage capabilities.  
 CC Sequences AAC60016-C60024 and AAB34853 are used in the course of the  
 CC invention during the identification and characterisation of the protein  
 CC and nucleotide sequences.  
 XX  
 SQ Sequence 97 AA;  
 XX

Query Match 41.9%; Score 496; DB 21; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-46;  
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 130 EDITEPQSIILAAAEKAGMSAEQAQGLLEKIATPKVKNKLTETEAACRYGAFGLPITVAH 189  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 1 editpeqsilaaekagmsaeqaglllekiatpkvknqlketteaacrygafglpitvah 60

Oy 190 VDGQTHMFGSDRMELLALHLLGCKWGPPIPPAVNARL 226  
 ||||||||||||||||||||||||||||||||||||||||||||

Db 61 vdgqthmfgsdrmellahlllgekwmgpippavnarl 97  
 ||||||||||||||||||||||||||||||||||||||||||||

RESULT 6  
 AAB34971  
 ID AAB34971 standard; Protein; 97 AA.  
 XX  
 AC AAB34971;  
 XX  
 DT 26-JAN-2001 (first entry)  
 XX  
 DE Gene 45 human secreted protein homologous amino acid sequence #175.  
 XX  
 KW Human; secreted protein; neuroprotective; cytostatic; cardioactive;  
 KW immunomodulatory; muscular; vulnary; gastrointestinal; nephrotropic;  
 KW antinefactive; gynaecological; antibacterial; neural disorder; cancer;  
 KW immune disease; reproductive disorder; proliferative disease;  
 KW gastrointestinal disease; wound healing; infectious disease;  
 KW food additive.  
 XX  
 OS Rattus sp.  
 XX  
 PN WO200056766-A1.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 16-MAR-2000; 2000WO-US06824.  
 XX  
 PR 19-MAR-1999; 99US-0125359.  
 XX  
 PD 03-DEC-1999; 99US-0168664.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 DR WPI: 2000-594574/56.  
 XX  
 PT Human secreted proteins and gene sequences encoding them, useful for  
 PT detection, prevention, and treatment of various disorders such as  
 PT cancer and immune system disorders -  
 XX  
 PS Disclosure; Page 83; 442pp; English.  
 XX  
 CC The polynucleotide sequences given in AAC60025-C60071 encode the human  
 CC secreted proteins represented in AAB34854-B34900. Sequences  
 CC AAB34901-B34976 are fragments of proteins encoded by the genes, and also  
 CC proteins with which they share sequence homology. The proteins have  
 CC activities based on the tissues in which their encoding genes are  
 CC expressed. Examples of the proteins activities include: neuroprotective;  
 CC - cytostatic; cardioactive; immunomodulatory; general muscular activity;  
 CC antinefactive; gynaecological; and antibacterial. The human secreted  
 CC proteins, polynucleotides, antagonists and antagonists of the invention  
 CC may be useful in treating, preventing and/or diagnosing various  
 CC diseases, disorders and conditions such as neural, immune, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal and  
 CC proliferative disorders and cancer. They may also be used in the  
 CC treatment of wounds, and infectious diseases. The polypeptides may be  
 CC used as a food additive or preservative to increase storage capabilities.  
 CC Sequences AAC60016-C60024 and AAB34853 are used in the course of the  
 CC invention during the identification and characterisation of the protein  
 CC and nucleotide sequences.  
 XX  
 SQ Sequence 97 AA;  
 XX

Query Match 32.7%; Score 387; DB 21; Length 97;  
 Best Local Similarity 76.3%; Pred. No. 1.4e-34;  
 Matches 74; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

us-09-441-723-1.1.rag

Tue May 21 08:04:00 2002

QY 130 EDITEPQSIILAAAEKAGMSAQGLLEKATPKVKNOLKETTAAACRYGAFGLPITVAH 189  
Db 1 editesqnllsaaekagmataqahllnkistelvkskirettgaackyygaiglpptvah 60  
QY 190 VDGQTHLFGSDRMELLAHLGKWMGPPIPPAVNARL 226  
Db 61 vdgktylmfsgdrmlaylllgkwmgyppptlnarl 97

RESULT 7  
AAW96474  
ID AAW96474 standard; Protein; 107 AA.  
XX \*  
AC AAW96474;  
DT 21-NOV-2001 (first entry)  
DE - Human reproductive system related antigen SEQ ID NO: 5132.  
KW Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy.  
XX Homo sapiens.  
OS WO200155320-A2.  
PN 02-AUG-2001.  
PD  
PE 17-JAN-2001; 2001WO-US01339.  
PF  
PG  
PH  
PI 31-JAN-2000; 2000US-0179065.  
PJ 04-FEB-2000; 2000US-0180628.  
PK 24-FEB-2000; 2000US-0184664.  
PL 02-MAR-2000; 2000US-0186350.  
PM 16-MAR-2000; 2000US-0189874.  
PN 17-MAR-2000; 2000US-0190076.  
PO 18-MAR-2000; 2000US-0198123.  
PP 19-MAY-2000; 2000US-0205515.  
PQ 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PS 07-JUL-2000; 2000US-0216647.  
PT 07-JUL-2000; 2000US-0216680.  
PU 11-JUL-2000; 2000US-0217487.  
PV 11-JUL-2000; 2000US-0217496.  
PW 14-JUL-2000; 2000US-0218290.  
PX 26-JUL-2000; 2000US-0220963.  
PY 26-JUL-2000; 2000US-0220964.  
PZ 14-AUG-2000; 2000US-0224518.  
QA 14-AUG-2000; 2000US-0224519.  
QB 14-AUG-2000; 2000US-0225213.  
QC 14-AUG-2000; 2000US-0225214.  
QD 14-AUG-2000; 2000US-0225266.  
QE 14-AUG-2000; 2000US-0225267.  
QF 14-AUG-2000; 2000US-0225268.  
QG 14-AUG-2000; 2000US-0225270.  
QH 14-AUG-2000; 2000US-0225275.  
QI 14-AUG-2000; 2000US-0225278.  
QJ 14-AUG-2000; 2000US-0225279.  
QK 18-AUG-2000; 2000US-0226279.  
QL 22-AUG-2000; 2000US-0226681.  
QM 22-AUG-2000; 2000US-0226686.  
QN 22-AUG-2000; 2000US-0227182.  
QO 23-AUG-2000; 2000US-0227009.  
QP 30-AUG-2000; 2000US-0228924.  
QR 01-SEP-2000; 2000US-0229287.  
QS 01-SEP-2000; 2000US-0229343.  
QT 01-SEP-2000; 2000US-0229344.  
QU 01-SEP-2000; 2000US-0229345.  
QV 05-SEP-2000; 2000US-0229509.  
QW 05-SEP-2000; 2000US-0229513.  
QX  
QY  
QZ  
RA  
RB  
RC  
RD  
RE  
RF  
RG  
RH  
RI  
RJ  
RK  
RL  
RM  
RN  
RO  
RP  
RQ  
RS  
RT  
RU  
RV  
RW  
RX  
RY  
RZ  
SA  
SB  
SC  
SD  
SE  
SF  
SG  
SH  
SI  
SJ  
SK  
SL  
SM  
SN  
SO  
SP  
SQ  
SR  
SS  
ST  
SU  
SV  
SW  
SX  
SY  
SZ  
TA  
TB  
TC  
TD  
TE  
TF  
TG  
TH  
TI  
TJ  
TK  
TL  
TM  
TN  
TO  
TP  
TQ  
TR  
TS  
TT  
TU  
TV  
TW  
TX  
TY  
TZ  
UA  
UB  
UC  
UD  
UE  
UF  
UG  
UH  
UI  
UJ  
UK  
UL  
UM  
UN  
UO  
UP  
UQ  
UR  
US  
UT  
UU  
UV  
UW  
UX  
UY  
UZ  
VA  
VB  
VC  
VD  
VE  
VF  
VG  
VH  
VI  
VJ  
VK  
VL  
VM  
VN  
VO  
VP  
VQ  
VR  
VS  
VT  
VU  
VV  
VV  
VW  
VX  
VY  
VZ  
WA  
WB  
WC  
WD  
WE  
WF  
WG  
WH  
WI  
WJ  
WK  
WL  
WM  
WN  
WO  
WP  
WQ  
WR  
WS  
WT  
WU  
WV  
WW  
WX  
WY  
WZ  
XA  
XB  
XC  
XD  
XE  
XF  
XG  
XH  
XI  
XJ  
XK  
XL  
XM  
XN  
XO  
XP  
XQ  
XR  
XS  
XT  
XU  
XV  
XW  
XX  
XY  
XZ  
YA  
YB  
YC  
YD  
YE  
YF  
YG  
YH  
YI  
YJ  
YK  
YL  
YM  
YN  
YO  
YP  
YQ  
YR  
YS  
YT  
YU  
YV  
YW  
YX  
YY  
YZ  
ZA  
ZB  
ZC  
ZD  
ZE  
ZF  
ZG  
ZH  
ZI  
ZJ  
ZK  
ZL  
ZM  
ZN  
ZO  
ZP  
ZQ  
ZR  
ZS  
ZT  
ZU  
ZV  
ZW  
ZX  
ZY  
ZZ

```

XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-6393362/73.
DR
XX N-PSDB; AAS90383.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID NO 56555; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation (I) and
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1086 AA;
SQ
Query Match
Best Local Similarity 7.68; Score 89.5; DB 22: Length 1086;
Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;
QY 126 WSRNEDITEPQSIILAAAEKAGMSAEQA-----OGLLEKIATPKVKNQ 167
Db 301 wecnkdvvvdkkqsvfekagntsdslaavfvptgydeytssayssllgqvfpvfyng 360
QY 168 LKETTETACRYAGFGLPITVAHVDDGQTHFLFGSDRMELLALHLGSKWMGPPIPA 221
Db 361 lr--leeqlgyavfafpmsvgrqwgmgflqsndkq---psflwerykaffpta 409
RESULT 9
AAY32947
ID AAY32947 standard; Protein; 545 AA.
XX
XX AAY32947;
AC
XX
XX 09-NOV-1999 (first entry)
DT
XX Mutant threonine dehydratase/deaminase protein sequence.
DE
XX Threonine dehydratase/deaminase; TD; feedback insensitive mutant;
KW molecular marker; isoleucine toxic structural analog resistance;
KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;
KW polyhydroxybutyrate; antibiotic resistance marker; mutagen
XX

```



Tue May 21 08:04:00 2002

Threonine dehydratase/deaminase; OMRI gene; feedback inhibition; transgenic plant; selectable marker; isoleucine; mutant.

XX Arabidopsis thaliana.  
KW Synthetic.  
XX WO9941395-A1.  
XX 19-AUG-1999.  
XX 08-JAN-1999; 99WO-US00560.  
XX 10-JUL-1998; 98WO-US14362.  
XX 17-FEB-1998; 98US-0074875.  
XX (DOMC) DOW AGROSCIENCES LLC.  
XX (PURD) PURDUE RES FOUND.  
XX Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
XX WPI; 1999-527375/44.  
XX -N-PSDB; AAZ11205.

Location/Qualifiers  
Key 439..457  
Region /note= "regulatory region R4"  
Region 489..507  
Region /note= "regulatory region R6"  
Misc-difference 452  
FT /note= "Arg in wild-type TD"  
FT Misc-difference 497  
FT /note= "Arg in wild-type TD"

WO9902656-A1.  
21-JAN-1999.  
10-JUL-1998; 98WO-US14362.  
17-FEB-1998; 98US-0074875.  
10-JUL-1997; 97US-0052096.  
(PURD) PURDUE RES FOUND.

Mourad GS;  
WPI; 1999-120860/10.  
N-PSDB; AAX25340.  
New sequences encode mutant threonine dehydratase/deaminase - which is insensitive to feedback inhibition, useful as a selective marker to produce transformed cells resistant to toxic isoleucine analogues  
Disclosure; Page 71-73; 120pp; English.

The present sequence represents an Arabidopsis thaliana mutant threonine dehydratase/deaminase (TD) protein which, unlike wild-type TD, is insensitive to feedback inhibition by isoleucine. Claimed polynucleotides (see AAX25332-40), originally isolated and cloned from A. thaliana mutated line GM1b (omri/omri1), encode feedback insensitive TD that can be used to transform a wide variety of plants, fungi, bacteria and yeast. Mutant TD differs from the wild-type enzyme only by an R499C amino acid substitution in the regulatory region R4, and by an R544H substitution in regulatory region R6. Mutant TD is not only insensitive to feedback inhibition by isoleucine, but is also insensitive to structural analogues of isoleucine that are toxic to plants and microorganisms which synthesize only wild-type TD. Nucleotide sequences encoding mutated forms of TD can therefore be used to create cells that are insensitive to compounds normally toxic to cells expressing only wild-type TD enzymes, and thus may be used to provide a biochemical selectable marker. Transformants that express the mutant TD show increased levels of isoleucine production, and thus provide an improved nutrient source.

XX Sequence 545 AA;  
Query Match 7.5%; Score 89; DB 20; Length 545;  
Best Local Similarity 22.8%; Pred. No. 0.95; 57; Indels 80; Gaps 11;  
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
QY 55 PPGL-LPRKGLYMANDLKLRHLLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
DB 5 PPKLIPR-----LKVSPNSLQ-----YPAGYLGAVPERTNEAENGSAEAMEYLTNI 52  
QY 105 -----AVNLEHP-EMLEKASRELWMRVWSNEDITE-----POSILAA 141  
DB 53 lstkvydialesplqlakkskrlgvrmvylkredlqpvfsklrgaynmvklpadqlak 112  
QY 142 AEKAGMSAEAOAG-----LLEKIATPKVKNQLKRTTEACRYGAFGLPITVA 188  
DB 113 gvicsagbnhagqvalasklgctavimpvttpeikwqavnl----- 156  
QY 189 HVDGQTHMFGSDRMELLALH-----LGEKMGPIPP 220  
DB 157 ---gatvlfgdsydaqahakiraeegltfipp 188

RESULT 10  
ID AAY05711  
XX AAY05711 standard; Protein; 545 AA.  
XX AAY05711;  
XX 19-JUL-1999 (first entry)  
XX Feedback insensitive threonine dehydratase/deaminase.

XX Arabidopsis thaliana.  
KW Synthetic.  
XX WO9941395-A1.  
XX 19-AUG-1999.  
XX 08-JAN-1999; 99WO-US00560.  
XX 10-JUL-1998; 98WO-US14362.  
XX 17-FEB-1998; 98US-0074875.  
XX (DOMC) DOW AGROSCIENCES LLC.  
XX (PURD) PURDUE RES FOUND.  
XX Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
XX WPI; 1999-527375/44.  
XX -N-PSDB; AAZ11205.  
XX New nucleic acid encoding threonine dehydratase deaminase resistant to feedback inhibition, useful as selection marker for cell transformation and to impart herbicide resistance  
XX Claim 13; Page 131-134; 194pp; English.  
XX This sequence represents a mutant Arabidopsis thaliana threonine dehydratase/deaminase (TD) protein of the invention. The protein is a feedback insensitive mutant. The TD DNA sequence is used as molecular marker (imparting resistance to toxic structural analogues of isoleucine) for selecting transformed cells and to produce transformants with increased levels of isoleucine (and thus better nutritional value) or of intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use of the isoleucine structural analogues as herbicides. The DNA sequences are alternatives for antibiotic resistance markers (which are potentially harmful to the environment). Since no human analog of TD exists (humans can not synthesize isoleucine), it should be safe to use.

XX Sequence 545 AA;  
Query Match 7.5%; Score 89; DB 20; Length 545;  
Best Local Similarity 22.8%; Pred. No. 0.95; 57; Indels 80; Gaps 11;  
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
QY 55 PPGL-LPRKGLYMANDLKLRHLLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
DB 5 PPKLIPR-----LKVSPNSLQ-----YPAGYLGAVPERTNEAENGSAEAMEYLTNI 52  
QY 105 -----AVNLEHP-EMLEKASRELWMRVWSNEDITE-----POSILAA 141  
DB 53 lstkvydialesplqlakkskrlgvrmvylkredlqpvfsklrgaynmvklpadqlak 112

QY 142 AERAGMSAEQAQ-----LLEKIATPKVKLNKLTETEAACRYGAFGLPTVA 188  
 Db 113 gvicssaghnagqvalasklgctavimpvttpeikwqavlenl----- 156  
 QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
 Db 157 ---gatvvlfqdsydaqahakiraeegltfipp 188

## RESULT 11

AAAY32950  
 ID AAY32950 standard; Protein: 590 AA.

AC AAY32950;

XX 09-NOV-1999 (first entry)

DE Mutant threonine dehydratase/deaminase protein sequence.

XX Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutein.

XX Arabidopsis thaliana.

OS Synthetic.

XX WO9941395-A1.

PN 19-AUG-1999.

PD 08-JAN-1999; 99WO-US00560.

PF 10-JUL-1998; 98WO-US14362.

XX 17-FEB-1998; 98US-0074875.

XX (DOWC) DOW AGROSCIENCES LLC.

PA (PURD) PURDUE RES FOUND.

XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;

PI WPI; 1999-527375/44.

DR N-PSDB; AA211208.

XX New nucleic acid encoding threonine dehydratase deaminase resistant

PT to feedback inhibition, useful as selection marker for cell

PT transformation and to impart herbicide resistance

XX Example 3; Page 119-123; 194pp; English.

PS This sequence represents a mutant Arabidopsis thaliana threonine

XX dehydratase/deaminase (TD) protein of the invention. The protein is a

CC feedback insensitive mutant. The TD DNA sequence is used as molecular

CC marker (imparting resistance to toxic structural analogues of isoleucine

CC for selecting transformed cells and to produce transformants with

CC increased levels of isoleucine (and thus better nutritional value) or of

CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for

CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also

CC TD-expressing plants permit use of the isoleucine structural analogues as

CC herbicides. The DNA sequences are alternatives for antibiotic resistance

CC markers (which are potentially harmful to the environment). Since no

CC human analog of TD exists (humans can not synthesize isoleucine), it

CC should be safe to use.

XX Sequence 590 AA;

QY Query Match

Best Local Similarity 7.5%; Score 89; DB 20; Length 590;

Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGCL-LPRKGLYMANDLKLRHHLQIPHFPPKDFLSVM-----LEKGSLS-AMRFLT-- 104

Db 50 ppklplpr-----lkvspnslq---ypagilqavpertenaeagsiaameyltnl 97  
 QY 105 -----AVNLEHP-EMLEKASRELWMRVWSRNEDETE-----PQSIIAA 141  
 Db 98 lstkvydiaiesplqiaakkskrigvrnylkrredlqpvsfklrgaynmvmklpadqlak 157  
 QY 142 AEKAGMSAEQAQ-----LLEKIATPKVKLNKLTETEAACRYGAFGLPTVA 188  
 Db 158 gvicssaghnagqvalasklgctavimpvttpeikwqavlenl----- 201  
 QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
 Db 202 ---gatvvlfqdsydaqahakiraeegltfipp 233

## RESULT 12

AAAY32939

ID AAY32939 standard; Protein: 592 AA.

AC AAY32939;

XX 09-NOV-1999 (first entry)

DE Mutant threonine dehydratase/deaminase protein sequence.

XX Threonine dehydratase/deaminase; TD; feedback insensitive mutant;

KW molecular marker; isoleucine toxic structural analog resistance;

KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;

KW polyhydroxybutyrate; antibiotic resistance marker; mutein.

XX Arabidopsis thaliana.

OS Synthetic.

XX WO9941395-A1.

PN 19-AUG-1999.

PD 08-JAN-1999; 99WO-US00560.

PF 10-JUL-1998; 98WO-US14362.

XX 17-FEB-1998; 98US-0074875.

XX (DOWC) DOW AGROSCIENCES LLC.

PA (PURD) PURDUE RES FOUND.

XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;

PI WPI; 1999-527375/44.

DR N-PSDB; AA211197.

XX New nucleic acid encoding threonine dehydratase deaminase resistant

PT to feedback inhibition, useful as selection marker for cell

PT transformation and to impart herbicide resistance

XX Claim 13; Fig 8; 194pp; English.

PS This sequence represents a mutant Arabidopsis thaliana threonine

CC dehydratase/deaminase (TD) protein of the invention. The protein is a

CC feedback insensitive mutant. The TD DNA sequence is used as molecular

CC marker (imparting resistance to toxic structural analogues of isoleucine

CC for selecting transformed cells and to produce transformants with

CC increased levels of isoleucine (and thus better nutritional value) or of

CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for

CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also

CC TD-expressing plants permit use of the isoleucine structural analogues as

CC herbicides. The DNA sequences are alternatives for antibiotic resistance

CC markers (which are potentially harmful to the environment). Since no

CC human analog of TD exists (humans can not synthesize isoleucine), it

CC should be safe to use.

XX Sequence 592 AA;

100

\_\_\_\_\_

CC to produce transformants with increased levels of isoleucine (and thus  
CC better nutritional value) or of intermediates in biosynthesis of  
CC isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable  
CC biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use  
CC of the isoleucine structural analogues as herbicides. The DNA sequences  
CC are alternatives for antibiotic resistance markers (which are potentially  
CC harmful to the environment). Since no human analog of TD exists (humans  
CC can not synthesize isoleucine), it should be safe to use.

XX Sequence 592 AA;

Query Match 7.5%; Score 89; DB 20; Length 592;  
Best Local Similarity 22.8%; Pred. No. 1.1;  
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHFPRKDFLSVM-----LEKGSLS-AMRFLT-- 104  
Db 52 ppklpplr-----lkvspnslq---ypagylgavpertenaeagsiaameyltni 99  
QY 105 -----AVNLEHP-EMLEKASRELMMRWVSRNEDITE-----POSILAA 141  
Db 100 lskvydiaiesplqlaklskrigvmlykredlqpvsfklrgaynmvklpadqlak 159  
QY 142 AEKAGMSAEQAQG-----LLEKIATPKVKNQKLTETEAACRYGAFGLPITVA 188  
Db 160 gvicssagnhagqvalsasklgtavimpvttpeikwqavenl----- 203  
QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
Db 204 ---gatvvifgdsydaqahakiraeegitfipp 235

RESULT 15

AAV05702  
ID AAY05702 standard; Protein; 592 AA.

XX AAY05702;

DT 19-JUL-1999 (first entry)

DE Arabidopsis wild-type threonine dehydratase/deaminase.

KW Threonine dehydratase/deaminase; OMRI gene; feedback inhibition;  
KW transgenic plant; selectable marker; isoleucine.  
XX Arabidopsis thaliana.

Key Location/Qualifiers  
FT Peptide 1..90  
FT Protein /note= "transit peptide"  
FT Region 91..592  
FT Region /note= "mature protein"  
FT Region 486..504  
FT Region /note= "regulatory region R4"  
FT Region 536..554  
FT Region /note= "regulatory region R6"

XX WO9902656-A1.

XX 21-JAN-1999.

XX 10-JUL-1998; 98WO-US14362.

XX 17-FEB-1998; 98US-0074875.

XX 10-JUL-1997; 97US-0052096.

XX (PURD ) PURDUE RES FOUND.

XX Mourad GS;

XX WPI; 1999-120860/10.

XX N-PSDB; AAX25331.

XX

PT New sequences encode mutant threonine dehydratase/deaminase - which  
PT is insensitive to feedback inhibition, useful as a selective marker  
PT to produce transformed cells resistant to toxic isoleucine analogues  
XX Disclosure; Page 50-53; 120pp; English.

XX The present sequence represents Arabidopsis thaliana var. Columbia  
CC wild-type threonine dehydratase/deaminase (TD), the first enzyme of  
CC the isoleucine biosynthetic pathway. TD is encoded by the OMRI  
CC gene (see AAX25331) of A. thaliana. The invention provides nucleotide  
CC sequences (see AAX25332-40), originally isolated and cloned from A.  
CC thaliana mutated line GMI1b (omri/omri), which encode feedback  
CC insensitive TD that can be used to transform a wide variety of  
CC plants, fungi, bacteria and yeast. The mutated form of TD differs  
CC from the wild-type only by 2 point mutations (C to T at nucleotide  
CC 1495, and G to A at nucleotide 1631), which result in an R499C amino  
CC acid substitution in the regulatory region R4 of TD, and an R544H  
CC substitution in the regulatory region R6. These forms of TD are not  
CC only insensitive to feedback inhibition by isoleucine, but are also  
CC plants and microorganisms which synthesize only wild-type TD.  
CC Nucleotide sequences encoding mutated forms of TD can therefore be  
CC used to create cells that are insensitive to compounds normally  
CC toxic to cells expressing only wild-type TD enzymes, and thus may  
CC be used to provide a biochemical selectable marker. Transformants  
CC harboring a nucleotide sequence comprising a promoter operably  
CC linked to a mutated TD sequence demonstrate increased levels of  
CC isoleucine production, and thus provide an improved nutrient source.

XX Sequence 592 AA;

Query Match 7.5%; Score 89; DB 20; Length 592;  
Best Local Similarity 22.8%; Pred. No. 1.1;  
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHFPRKDFLSVM-----LEKGSLS-AMRFLT-- 104  
Db 52 ppklpplr-----lkvspnslq---ypagylgavpertenaeagsiaameyltni 99

QY 105 -----AVNLEHP-EMLEKASRELMMRWVSRNEDITE-----POSILAA 141  
Db 100 lskvydiaiesplqlaklskrigvmlykredlqpvsfklrgaynmvklpadqlak 159

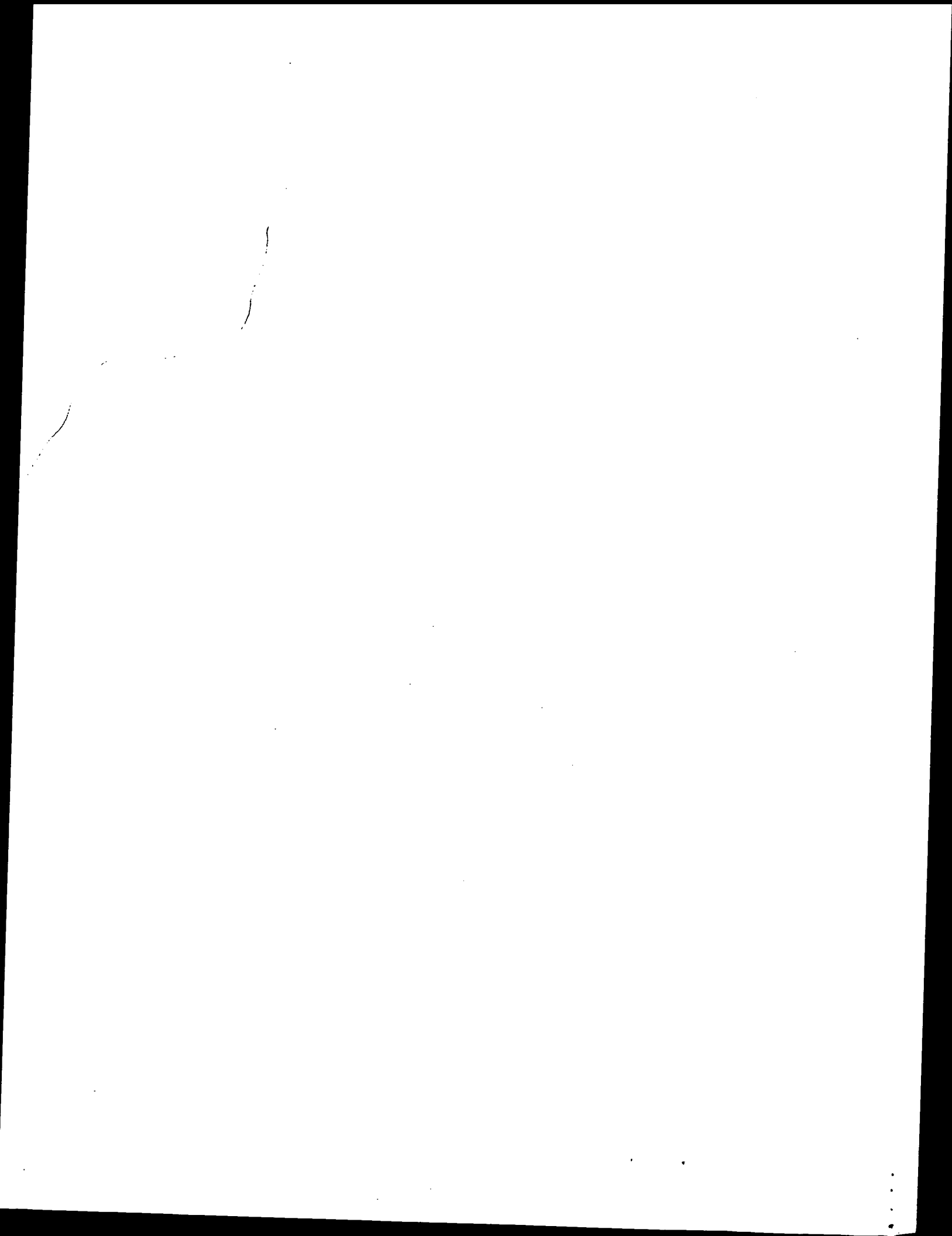
QY 142 AEKAGMSAEQAQG-----LLEKIATPKVKNQKLTETEAACRYGAFGLPITVA 188  
Db 160 gvicssagnhagqvalsasklgtavimpvttpeikwqavenl----- 203

QY 189 HVDGQTHMLFGSDRMELLAHL---LGEKWMGPPIP 220  
Db 204 ---gatvvifgdsydaqahakiraeegitfipp 235

Search completed: May 20, 2002, 08:39:12  
Job time: 344 sec

us-09-441-723-1.rag

Tue May 21 08:04:00 2002



us-09-441-723-1.ra1

Tue May 21 08:04:01 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:33:48 ; Search time 13.17 Seconds  
(without alignments)  
419.148 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 1185  
Sequence: 1 MGPLPRTVELFDVLSY.....AHLGKMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution..

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1185	100.0	226	3	US-08-978-174-1
2	874	73.8	226	3	US-08-978-174-3
3	80.5	6.8	339	2	US-08-855-714-3
4	76.5	6.5	317	2	US-09-066-075-2
5	76.5	6.5	317	2	US-08-518-615A-2
6	76.5	6.5	317	3	US-08-951-889-2
7	76.5	6.5	317	4	US-09-472-857-2
8	76.5	6.5	497	1	US-08-075-193-4
9	76.5	6.5	497	2	US-08-564-090A-4
10	76.5	6.5	497	5	PCT-US94-06698-4
11	75.5	6.4	1398	1	US-08-750-532-9
12	75.5	6.4	1398	4	US-08-894-818B-8
13	75.5	6.4	1398	4	US-09-445-472-6
14	73.5	6.2	523	2	US-08-473-553A-3
15	73.5	6.2	869	1	US-08-188-582-32
16	73.5	6.2	869	1	US-08-646-715-32
17	73.5	6.2	980	2	US-08-473-553A-6
18	73.5	6.2	985	2	US-08-473-553A-2
19	73.5	6.2	948	1	US-08-698-551-14
20	73.5	6.2	948	2	US-08-602-228-14
21	73.5	6.2	948	2	US-08-533-901B-14
22	73.5	6.2	948	2	US-08-839-032A-14
23	73.5	6.2	948	2	US-08-839-031A-14
24	73.5	6.2	948	5	PCT-US95-12724-14
25	71.5	6.0	461	4	US-09-346-408-8
26	71.5	6.0	532	4	US-09-181-336-15
27	71.5	6.0	543	2	US-08-922-170B-10

28	71.5	6.0	543	4	US-09-071-739B-2
29	71.5	6.0	543	4	US-09-181-336-13
30	71.5	6.0	543	4	US-09-260-038B-2
31	71.5	6.0	587	1	US-07-935-905A-23
32	71.5	6.0	1090	4	US-09-346-237-5
33	70.5	5.9	615	2	US-08-484-101B-38
34	70.5	5.9	615	4	US-08-714-524D-38
35	70.5	5.9	615	4	US-08-198-446B-11
36	70.5	5.9	401	1	US-08-870-693-11
37	69.5	5.9	259	2	US-07-857-224B-51
38	69.5	5.9	529	4	US-08-426-509A-15
39	69.5	5.9	529	5	PCT-US95-05008-15
40	69.5	5.9	1075	1	US-07-623-033-2
41	69.5	5.8	689	1	US-07-766-351-5
42	69.5	5.8	689	1	US-08-059-032-5
43	69.5	5.8	689	5	PCT-US91-07290-5
44	68.5	5.8	339	3	US-08-758-280-1
45	68.5	5.8	339	3	US-08-758-280-2

## ALIGNMENTS

RESULT 1  
US-08-978-174-1  
; Sequence 1, Application US/08978174  
; Patent No. 6030809  
; GENERAL INFORMATION:  
; APPLICANT: Shah, Purvi  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ For Windows Version 2.0  
; CURRENT APPLICATION NUMBER: US/08/978,174  
; APPLICATION NUMBER: US/08/978,174  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0430 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 226 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BLADTUT04  
; CLONE: 1554593  
US-08-978-174-1

	Query Match	100.0%;	Score 1185;	DB 3;	Length 226;
	Best Local Similarity	100.0%;	Pred. No. 5.3e-126;		
	Matches 226;	Conservative	0;	Mismatches 0;	Indels 0;
	Gaps				
QY	1	MGPLPRTVELFYDVLSPSYMWLGFEILCRYQNIWNINLQLRPSLITGIMKDSGNKPPGLLP	60		
Db	1	MGPLPRTVELFYDVLSPSYMWLGFEILCRYQNIWNINLQLRPSLITGIMKDSGNKPPGLLP	60		
QY	61	RKGLYMANDLKLRHHLQIPIHFPKDFLSVMLEKGLSAMRELTAVNLEHPENMLEKASRE	120		
Db	61	RKGLYMANDLKLRHHLQIPIHFPKDFLSVMLEKGLSAMRELTAVNLEHPENMLEKASRE	120		
QY	121	LWMRVMSRNEDITEPOSILAAAEKAGMSAQOGLLEKIATPKVKNQLKETTEAAACRYGA	180		
Db	121	LWMRVMSRNEDITEPOSILAAAEKAGMSAQOGLLEKIATPKVKNQLKETTEAAACRYGA	180		
QY	181	FGLPITVAHVDCQTHMLFGSDRMELLAHLGKWMGPIPPAVNARL	226		
Db	181	FGLPITVAHVDCQTHMLFGSDRMELLAHLGKWMGPIPPAVNARL	226		

RESULT# 3  
 US-08-855-714-3  
 ; Sequence 3, Application US/08855714  
 ; Patent No. 5939075  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Houng, Huo-Shu H.  
 ; APPLICANT: Warren, Richard L.  
 ; TITLE OF INVENTION: MUTANTS OF BRUCELLA MELITENSIS  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: John Moran, Esq.  
 ; STREET: HQ USMRDC, Dept. of Army, Fort Detrick  
 ; CITY: Frederick  
 ; STATE: MD  
 ; COUNTRY: US  
 ; ZIP: 21702-5012  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/855,714  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/334,129  
 ; FILING DATE: 04-NOV-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hendricks, Glenna  
 ; REGISTRATION NUMBER: 32,535  
 ; REFERENCE/DOCKET NUMBER: 08/143,692  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 619-2065  
 ; TELEFAX: (301) 619-7714  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 339 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: both  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE:  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: BRUCELLA MELITENSIS  
 ; IS-08-855-714-3



us-09-441-723-1-rai

Tue May 21 08:04:01 2002

QY 117 ASRE-----LWVRVWSRNE-----ITEPQS----- 137  
 Db 103 DPEHKEFRLALWKQIADRYKDYPTLFEILNEPHGNLTPEKNWELLEALKVIRSIDK 162  
 QY 138 -----ILAAAEKAGSAEQAGLLEKIAIATPKVKYNQKLETTAEACRYCAFGLPITVAHVDGQ 193  
 Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE-----KNSIVTIHYINPFEF----- 203  
 QY 194 THMLFGSDRMELLALHLLGCKWMP 217  
 Db 204 THQ--GAEWVEGSEKWLGRKWCSP 225

RESULT 5  
 US-08-518-615A-2  
 ; Sequence 2, Application US/08518615A  
 ; Patent No. 5962258  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mathur, E., et al.  
 ; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
 ; STREET: 6 BECKER FARM ROAD  
 ; CITY: ROSELAND  
 ; STATE: NEW JERSEY  
 ; COUNTRY: USA  
 ; ZIP: 07068  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 INCH DISKETTE  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: WORD PERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/518,615A  
 ; FILING DATE: August 23, 1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: FERRARO, GREGORY D.  
 ; REGISTRATION NUMBER: 36,134  
 ; REFERENCE/DOCKET NUMBER: 331400-20  
 ; TELEPHONE: 201-994-1700  
 ; TELEFAX: 201-994-1744  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 317 AMINO ACIDS  
 ; TYPE: AMINO ACID  
 ; STRANDEDNESS:  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: PROTEIN  
 ; US-08-518-615A-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
 Best Local Similarity 21.1%; Pred. No. 2;  
 Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
 QY 76 HLQIPH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPMELEK 116  
 Db 49 HVRIPRWSTHAYAFPPYKIMDRFRRVDEVINGALKRG-----LAVAINIHYYNPF 102  
 QY 117 ASRE-----LWVRVWSRNE-----ITEPQS----- 137  
 Db 103 DPEHKEFRLALWKQIADRYKDYPTLFEILNEPHGNLTPEKNWELLEALKVIRSIDK 162  
 QY 138 -----ILAAAEKAGSAEQAGLLEKIAIATPKVKYNQKLETTAEACRYCAFGLPITVAHVDGQ 193  
 Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE-----KNSIVTIHYINPFEF----- 203

Db 47 PLPTSPMNLKWCQSAPPTSLKRRFLFCPPPHWK-----SLRTASOKSSFSTKAALK 99  
 QY 60 PRK-GLYMANDKLLRHHLIQIPIHFKDFLSVMLEKGS-----SAMRFLTAVNLEH 110  
 Db 100 PRPGSWMTKRKSPRSALGRAWH-PODRRLGVYDGGQVRLASLDETQACNAFAIN-KA 157  
 QY 111 PEMLE---KASRELWVRVWSRNEITEPQSILAAAEKAG----- 146  
 Db 158 PAILEGFVEFREV-----SVIAARDRSGNVAIFDLAENVHDKGILATST 202  
 QY 147 -----MSAEQAQGLEKTA-----TPKVKYNQKLETT 173  
 Db 203 VPAAISVQTAAEAARTAEKILHALDYGVGLGFEFVLKDGTLLANEFAPRVHNS-GHWTE 261  
 QY 174 ACRYGAF-----GLPI--TVAHVDGQTHMLFGSD-----RMELLAHLGK 213  
 Db 262 ACAISQFEQHRAVAGLPLGNTDRHSDCVNENLIGDDIEKVPAILCEKNAVHLGYGK 320

RESULT 4  
 US-09-066-075-2  
 ; Sequence 2, Application US/09066075  
 ; Patent No. 5925749  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mathur, E., et al.  
 ; TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
 ; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
 ; STREET: 6 BECKER FARM ROAD  
 ; CITY: ROSELAND  
 ; STATE: NEW JERSEY  
 ; COUNTRY: USA  
 ; ZIP: 07068  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 INCH DISKETTE  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: WORD PERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/066,075  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/518,615  
 ; FILING DATE: August 23, 1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: FERRARO, GREGORY D.  
 ; REGISTRATION NUMBER: 36,134  
 ; REFERENCE/DOCKET NUMBER: 331400-20  
 ; TELEPHONE: 201-994-1700  
 ; TELEFAX: 201-994-1744  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 317 AMINO ACIDS  
 ; TYPE: AMINO ACID  
 ; STRANDEDNESS:  
 ; TOPOLOGY: LINEAR  
 ; MOLECULE TYPE: PROTEIN  
 ; US-09-066-075-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
 Best Local Similarity 21.1%; Pred. No. 2;  
 Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
 QY 76 HLQIPH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPMELEK 116  
 Db 49 HVRIPRWSTHAYAFPPYKIMDRFRRVDEVINGALKRG-----LAVAINIHYYNPF 102

QY 194 THMLFGSDRMELLALLGKWMGP 217  
||| : : : ||| : : :  
Db 204 THQ--GAEWEGSEKWLGRKWGSP 225

RESULT 6  
US-08-951-889-2  
; Sequence 2, Application US/08951889  
; Patent No. 6008032  
; GENERAL INFORMATION:  
; APPLICANT: Mathur, E., et al.  
; TITLE OF INVENTION: Carboxymethyl Cellulase from  
; TITLE OF INVENTION: Thermotoga Maritima  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/951,889  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/518,615  
; FILING DATE: August 23, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 331400-20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 317 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-951-889-2

Query Match 6.5%; Score 76.5; DB 3; Length 317;  
Best Local Similarity 21.1%; Pred. No. 2;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
QY 76 HLQPIH-----FP-----KDFLSVMLEKSGLSAMRFLTAVNLEHPMLEK 116  
||| : : : ||| : : :  
Db 49 HVRIPIRWSTHAYAPPYKIMDRFFKRVDEVINGALKRG-----LAVAINIHVEELMN 102  
QY 117 ASRE-----LWMRVWSRNE-----ITEPOS----- 137  
Db 103 DPEHKERFLALWKQIADRYKDYPTLFFELNEPHGNLTPEKNELLEALKVIRSIDK 162  
QY 138 -----ILAAAEKAGSAEQAGLLEKIATPKVNQKLTETEAACRYGAGFLPITVAHVDGQ 193  
Db 163 KHTIIGTAEMGGISA-----LEKLSVPKWE---KNSIVTIHYNPFEP----- 203  
QY 194 THMLFGSDRMELLALLGKWMGP 217  
||| : : : ||| : : :  
Db 204 THQ--GAEWEGSEKWLGRKWGSP 225

RESULT 8  
US-08-075-193-4  
; Sequence 4, Application US/08075193  
; Patent No. 5547868

RESULT 7  
US-09-472-857-2  
; Sequence 2, Application US/09472857  
; Patent No. 6245547  
; GENERAL INFORMATION:  
; APPLICANT: Mathur, E., et al.  
; TITLE OF INVENTION: Carboxymethyl Cellulase from  
; TITLE OF INVENTION: Thermotoga Maritima  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/472,857  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/951,889  
; FILING DATE:  
; APPLICATION NUMBER: 08/518,615  
; FILING DATE: August 23, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 331400-20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 317 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-09-472-857-2

Query Match 6.5%; Score 76.5; DB 4; Length 317;  
Best Local Similarity 21.1%; Pred. No. 2;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
QY 76 HLQPIH-----FP-----KDFLSVMLEKSGLSAMRFLTAVNLEHPMLEK 116  
||| : : : ||| : : :  
Db 49 HVRIPIRWSTHAYAPPYKIMDRFFKRVDEVINGALKRG-----LAVAINIHVEELMN 102  
QY 117 ASRE-----LWMRVWSRNE-----ITEPOS----- 137  
Db 103 DPEHKERFLALWKQIADRYKDYPTLFFELNEPHGNLTPEKNELLEALKVIRSIDK 162  
QY 138 -----ILAAAEKAGSAEQAGLLEKIATPKVNQKLTETEAACRYGAGFLPITVAHVDGQ 193  
Db 163 KHTIIGTAEMGGISA-----LEKLSVPKWE---KNSIVTIHYNPFEP----- 203  
QY 194 THMLFGSDRMELLALLGKWMGP 217  
||| : : : ||| : : :  
Db 204 THQ--GAEWEGSEKWLGRKWGSP 225

RESULT 8  
US-08-075-193-4  
; Sequence 4, Application US/08075193  
; Patent No. 5547868

COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatenIn Release #1.0, version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/564.090A  
FILING DATE: 02/05/96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD L. NEELEY, PH.D.  
REGISTRATION NUMBER: 30,092  
REFERENCE/DOCKET NUMBER: UCAL-236/01US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-843-5000  
TELEFAX: 415-857-0663  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-564-090A-4

	Query Match	6.5%;	Score 76.5;	DB 2;	Length 497;
	Best Local Similarity	25.4%;	Pred. No. 3.9;		
	Matches 44;	Conservative 29;	Mismatches 69;	Indels 31;	Gaps 9;
QY	14	VLSPYSLGFEILCRYONTWININQLRPSLITGIMKOSGNKPGPLPKRG-LYMANDLKL	72		
Db	195	LLTPPEHLEALLLCQRTDITKAAL-----GVLRQSRVKTVMVLRGRGPPLOVAFTIKE	246		
QY	73	LRHHLQI-----PIHEPKDFL-----SVMLEKGSLSAMRFUTAVNLEHPEMLEKASREL	121		
Db	247	LRMIQILPGARPILDVPDFLQDQIKVPRPRKRULTTELLRTAT--EKPGPAEARQAS	304		
QY	122	WMRYVWSRNEDETEQOSILAAAEKAGMSAEQAQGLLEKATPKVKNQOLKETTEA	174		
Db	305	ASRAWGL--RFRSPQQVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA	348		

RESULT 10  
PCT-US94-06698-4  
; Sequence 4, Application PC/TUS9406698  
; GENERAL INFORMATION:  
; APPLICANT: MILLER, WALTER L.  
; APPLICANT: HARIKRISHNA, JENNIFER A.  
; APPLICANT: BLACK, STEPHEN M.  
; TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ROBBINS, BERLINER & CARSON  
; STREET: 201 NORTH FIGUEROA STREET  
; CITY: LOS ANGELES  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 90012  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: PCT/US94/06698  
; APPLICATION NUMBER:  
; FILING DATE: FILED HERewith  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BERLINER, ROBERT  
; REGISTRATION NUMBER: 20,121  
; REFERENCE/DOCKET NUMBER: 5555-224-C1

GENERAL INFORMATION:  
APPLICANT: MILLER, WALTER L.  
APPLICANT: HARIKRISHNA, JENNIFER A.  
APPLICANT: BLACK, STEPHEN M.  
TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSER: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
STREET: FIVE PALO ALTO SQUARE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentLin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/075.193  
FILING DATE: 09-JUN-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: NEELEY Ph.D., RICHARD L.

```

REFERENCE/DOCKET NUMBER: UCAL-236/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-075-193-4

Query Match          6.5%      Score 76.5; DB 1; Length 497;
Best Local Similarity 25.4%;
Pred. No. 3.9;

```

```

QY 14 VLSPYSWLGFELICRQNIWNINQLRPSLITGIMKDSGNKPPGGLPRKG-LYMANDKL 72
Db 195 LLTPPEHLEALLCQRTDITKAAL-----GVLRQSRVKTWLVGRGPGLOVAFTIKE 246
QY 73 LRHHLOI-----PIHFPKDFL-----SVMLEKGLSAMRELTAVNLSEHPEMKASREL 121
Db 247 LREMIOLGARPILDVPDFGLQDKIKEVPRPKRLTELLRTAT--EKPGPAEAARQAS 304
QY 122 WMVWVSNRNIDITEPQSILAAAEKAGMSAEQAQGLLEKTATPKVKNLKETTETA 174
Db 305 ASRAWGL-RFRSPQQVLPSD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 9
US-08-564-090A-4
; Sequence 4, Application US/08564090A
; Patent No. 5939318
; Patent No. 5939318 5741703
; GENERAL INFORMATION:
; APPLICANT: MILLER, WALTER L.
; APPLICANT: HARIKRISHNA, JENNIFER A.
; APPLICANT: BLACK, STEPHEN M.
; TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD LLP
; STREET: FIVE PALO ALTO SQUARE
; STREET: 3000 EL CAMINO REAL
; CITY: PALO ALTO
; STATE: CALIFORNIA
;

```

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 213-977-1001  
 TELEFAX: 213-977-1003  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 497 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US94-06698-4

Query Match 6.5%; Score 76.5; DB 5; Length 497;  
 Best Local Similarity 25.4%; Pred. No. 3.9;  
 Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;  
 QY 14 VLSYSWGLGFEILCRQNTWNLQLRPSLITGIMKDSGNKPPGLPRKG-LYMANDLKL 72  
 Db 195 LITPPEHLEALLCQRTDITKAAL-----GVLRSRVKTVWLVRGRGLQVAFITKE 246  
 QY 73 LRHHLOI-----PIHPKDFL-----SVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121  
 Db 247 LREMIQPGARFILDVDFGLQDKIKEVPRPRKRLTELLRTAT--EKPGPAEAARQAS 304  
 QY 122 WMRYVSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIAIATPKVKNQKLTETEA 174  
 Db 305 ASRAWGL-REFRSPQVLFSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 11  
 US-08-750-532-9  
 ; Sequence 9, Application US/08750532  
 ; Patent No. 5756339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MITTA, Masanori  
 ; APPLICANT: YAMAMOTO, Katsuhiko  
 ; APPLICANT: MORISHITA, Mio  
 ; APPLICANT: ASADA, Kiyozo  
 ; APPLICANT: TSUNASAWA, Susumu  
 ; APPLICANT: KATO, Ikunoshin  
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE  
 ; NUMBER OF SEQUENCES: 18  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
 ; STREET: 419 Seventh Street N.W., Suite 300  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: United States of America  
 ; ZIP: 20004  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/750,532  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/JP95/01095  
 ; FILING DATE: 05-JUN-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 1994/130236  
 ; FILING DATE: 13-JUN-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 1994/173912  
 ; FILING DATE: 26-JUL-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BROWDY, Roger L.  
 ; REGISTRATION NUMBER: 25,618  
 ; REFERENCE/DOCKET NUMBER: MITTA-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1398 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-750-532-9

Query Match 6.4%; Score 75.5; DB 1; Length 1398;  
 Best Local Similarity 16.8%; Pred. No. 24;  
 Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;  
 QY 50 DSGNKPGLLPKGLY--MANDLKLRRHLOI----- 79  
 Db 1097 DKADFAVGUTPAEGLGEARNYTLIVKHALTLPEVPNATVIIGNVTYLTDEGNTVTFTYA 1156  
 QY 80 PIHPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRELWVRVSRNEDITEPOSIL 139  
 Db 1157 PTKLGSDEITVIVKKNFNTLEKTFQITVSEPEITE-----EDINEPKLAM 1202  
 QY 140 AAEEKAG--MSAE-QAQLLEKIAI-----PKVKNQKLTETEAACR 177  
 Db 1203 SSPKANATVSVMESEGGVKKTVTVEITINGTANETATVVPVPPKKAENIEVSDHVIS 1262  
 QY 178 YG-----AFGLPITVAHVDSGTHMLFGSDRMELLA-HILGEKW 214  
 Db 1263 YSIEGEVAKYVIITVKFASPVTVT---VTYIYAGPRVSILTLNFGYSW 1310

RESULT 12  
 US-08-894-818B-8  
 ; Sequence 8, Application US/08894818B  
 ; Patent No. 6261822  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TAKAKURA, Hikaru  
 ; APPLICANT: MORISHITA, Mio  
 ; APPLICANT: YAMAMOTO, Katsuhiko  
 ; APPLICANT: MITTA, Masanori  
 ; APPLICANT: ASADA, Kiyozo  
 ; APPLICANT: TSUNASAWA, Susumu  
 ; APPLICANT: KATO, Ikunoshin  
 ; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Browdy and Neimark  
 ; STREET: 419 Seventh Street N.W., Ste. 300  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: United States of America  
 ; ZIP: 20004  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/894,818B  
 ; FILING DATE: 20-MAY-1998  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/JP96/03253  
 ; FILING DATE: 07-NOV-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 323285/1995  
 ; FILING DATE: 12-DEC-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Browdy, Roger L.  
 ; REGISTRATION NUMBER: 25,618  
 ; REFERENCE/DOCKET NUMBER: TAKAKURA-1  
 ; TELECOMMUNICATION INFORMATION:

us-09-441-723-1-rai

Tue May 21 08:04:01 2002

TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1398 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-894-818B-8

Query Match 6.4%; Score 75.5; DB 4; Length 1398;  
Best Local Similarity 16.8%; Pred. No. 24;  
Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;  
QY 50 DSGNKPGLPRGLY--MANDLKLRLHQLI-----79  
Db 1097 DKADFAVGLTPAEGVLGEARNYTLIVKHALTLEPVPNATVIIGNYTLTDENGTVFTYA 1156  
QY 80 PIHPKDFLSVLMKSGLSAMRFLTAVNLEHPEMLEKASRELWMRVWSNEDITEPQSIL 139  
Db 1157 PTKGSDIIVIVKKNFTLEKTFQITVSEPEITE-----EDINEPKLAM 1202  
QY 140 AAARKAG--MSAE-OAQLLEKIAT-----PKVKNQKLTTEAAACR 177  
Db 1203 SSPEANATIVSMESEGGVKKTVTVEITINGTANETATIVVPVPPKKAENIEVSGDHVIS 1262  
QY 178 YG-----AFGLPITVAHVQDQTHLMFGSDRMELIA--HLLGEKW 214  
Db 1263 YSIEGEYAKYIIVTKFASPTVT-----VTYIVAGPRVSILTLNPLGYSW 1310

RESULT 13  
US-09-445-472-6  
Sequence 6, Application US/09445472  
Patent No. 6358726  
GENERAL INFORMATION:  
APPLICANT: TAKAKURA, HIKARU  
APPLICANT: MORISHITA, MIO  
APPLICANT: SHIMOJO, TOMOKO  
APPLICANT: ASADA, KIYOZO  
APPLICANT: KATO, IKUNOSHIN  
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
FILE REFERENCE: TAKAKURA-6  
CURRENT APPLICATION NUMBER: US/09/445,472  
CURRENT FILING DATE: 1999-12-06  
PRIOR APPLICATION NUMBER: 151969/1997  
PRIOR FILING DATE: 1997-06-10  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patent in version 3.0  
SEQ ID NO 6  
LENGTH: 1398  
TYPE: PRT  
ORGANISM: Pyrococcus furiosus  
US-09-445-472-6  
Query Match 6.4%; Score 75.5; DB 4; Length 1398;  
Best Local Similarity 16.8%; Pred. No. 24;  
Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;  
QY 50 DSGNKPGLPRGLY--MANDLKLRLHQLI-----79  
Db 1097 DKADFAVGLTPAEGVLGEARNYTLIVKHALTLEPVPNATVIIGNYTLTDENGTVFTYA 1156  
QY 80 PIHPKDFLSVLMKSGLSAMRFLTAVNLEHPEMLEKASRELWMRVWSNEDITEPQSIL 139  
Db 1157 PTKGSDIIVIVKKNFTLEKTFQITVSEPEITE-----EDINEPKLAM 1202  
QY 140 AAARKAG--MSAE-OAQLLEKIAT-----PKVKNQKLTTEAAACR 177  
Db 1203 SSPEANATIVSMESEGGVKKTVTVEITINGTANETATIVVPVPPKKAENIEVSGDHVIS 1262

QY 178 YG-----AFGLPITVAHVQDQTHLMFGSDRMELIA--HLLGEKW 214  
Db 1263 YSIEGEYAKYIIVTKFASPTVT-----VTYIVAGPRVSILTLNPLGYSW 1310  
RESULT 14  
US-08-473-553A-3  
Sequence 3, Application US/08473553A  
Patent No. 5859338  
GENERAL INFORMATION:  
APPLICANT: Meyerowitz, Elliot M.  
APPLICANT: Clark, Steven E.  
APPLICANT: Williams, Robert W.  
TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
TITLE OF INVENTION: Transformed Plants, and Proteins  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,553A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 523 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-473-553A-3

Query Match 6.2%; Score 73.5; DB 2; Length 523;  
Best Local Similarity 21.8%; Pred. No. 9.2;  
Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 14;  
QY 36 NLQRLPSLITGIMKDSGNKPP---GLLPKGLYMANDKLRLHQLIPIHPKDFLSVML 92  
Db 194 NLKHLHLFLHNNLTGHIPPELSGLVSLKSL---DLSI---NOLTGIPQSFINL-- 243  
QY 93 EKSLSAMRFLTAVNL-----EHPMLEKASRELWMRVWSNEDITEPQSILAAAE-- 143  
Db 244 --GN-----ITLINFRNLYGQIPEAIGELPKLEVFEVWENNFTLQLPANLGRNGLI 295  
QY 144 KAGMSAEQAQGLL-----EKIATPK-----VKNQKLET 171  
Db 296 KLDVSDNHLTGILPKDLCRGEKLEMLILSNFFGPTPEELGKCKSLTKIRIVKNLLNGT 355  
QY 172 TEACRYGAFGLP-ITVAHV-----DQTHLMFGSDRMELIAHLLGEKW--GP1PPAV 222  
Db 356 VPA---GLFNLPLVTIIELTDNFFSGELPVTMTSGDVLDOI--YLSNNWFSGEIPPAI 407  
RESULT 15

US-08-188-582-32  
; Sequence 32, Application US/08188582  
; Patent No. 5534410  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Nako  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,582  
; FILING DATE: 28-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 869 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-188-582-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;  
Best Local Similarity 20.5%; Pred. No. 20;  
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;  
QY 32 IWNNINQLRPSLITGIMKDSGNKPPGLLPKGLYMANDLKLRRHHHQIPIHFPPKDFLSVM 91  
Db 318 VWKFGKQMPQLLOAQVKEKGATGTSLSPP---HLPGLAICRSRGAVCLWSPEDGLRQI 373  
QY 92 LEKGSLSAMRELTAVNLEHPEMLKASRELWMLVWRSNEDITEPQSIILAAAEKAGMSAEQ 151  
Db 374 -----YRDPETLVFRDSSWR--WA---DFTAHPRVLTGDRGTGVKMLD 412  
QY 152 AQQ-----LLEKATPKVKQKLTETEAACRYGA-----FG-----LPITVAHVVG 192  
Db 413 TQGPFGGGLLLFRLG-----AEASCOKGERVLLTQYLGHSSPKCLPPTLHLVCT 461  
QY 193 QTHMFGSDRMELLALHLLGKWMGPPI-PAVNARL 226  
Db 462 QFSLYLVDERLPVPMI---KWNHGLPSPLLLARL 493

Search completed: May 20, 2002, 08:39:32  
Job time: 344 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2002, 08:34:43 ; Search time 19.29 Seconds  
(without alignments)  
1125.774 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 1185  
Sequence: 1 MGPLPRTVELFDVLSYVW.....AHLLEGKMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	328.5	27.7	226	2 T27747	hypothetical prote
2	298.5	25.2	225	2 T34201	hypothetical prote
3	198	16.7	195	2 G83629	hypothetical prote
4	164.5	13.9	194	2 S72164	2-hydroxychromene-
5	147.5	12.4	206	2 H87328	conserved hypotet
6	145.5	12.3	312	2 B88492	protein T07E3.3 li
7	140	11.8	197	2 T31286	2-nitrotoluene dio
8	125	10.5	33	2 S17164	glutathione transf
9	124	10.5	197	2 AB3212	2-hydroxychromene-
10	114	9.6	253	2 G87394	hypothetical prote
11	95	8.0	203	2 C55552	2-hydroxychromene-
12	91	7.7	581	2 H72425	ABC transporter, A
13	90.5	7.6	410	2 C69127	flavoprotein A hom
14	90.5	7.6	600	2 D95899	probable dipeptide
15	89.5	7.6	962	1 SNECPI	pitriylsin (EC 3.4
16	89.5	7.6	962	2 F91088	proteinase III [im
17	89.5	7.6	962	2 H85933	proteinase III [im
18	89	7.5	199	2 I49343	probable isomerase
19	89	7.5	592	2 T51712	threonine dehydrat
20	88.5	7.5	613	2 A35296	secretogranin II p
21	87.5	7.4	962	2 AC0865	protease III precu
22	87	7.3	926	2 E83375	probable glycosyl
23	85.5	7.2	224	2 A13595	frne protein (impo
24	84.5	7.1	316	2 B71301	probable tRNA delt
25	84.5	7.1	971	2 S54595	probable membrane
26	83.5	7.0	488	2 A40367	catalase (EC 1.11.
27	83.5	7.0	700	1 S09748	kinesin-related pr
28	83.5	7.0	1007	2 C84668	probable receptor-
29	82.5	7.0	263	2 T51169	hypothetical prote

frne protein VCA01  
polyketide biosynt  
secretogranin II -  
DNA-directed DNA p  
serine/threonine k  
hypothetical prote  
protein kinase CDC  
hypothetical prote  
topoisomerase I -  
catalase (EC 1.11.  
heat-stable entero  
rRNA methylase Spo  
ABC transporter (A  
hypothetical prote  
conserved hypotet  
alanine dehydrogen

ALIGNMENTS

RESULT 1  
T27747  
hypothetical protein ZK1320.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27747  
R:Berks, M.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z20414  
A:Accession: T27747  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-226 <WILL>  
A:Cross-references: EMBL:Z46934; PIDN:CAA87039.1; GSPDB:GN00020; CESP:ZK1320.1  
A:Experimental source: clone ZK1320  
C:Genetics:  
A:Gene: CESP:ZK1320.1  
A:Map position: 2  
A:Introns: 23/3; 177/3

Query Match 27.7%; Score 328.5; DB 2; Length 226;  
Best Local Similarity 33.0%; Pred. No. 7.2e-21;  
Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;  
QY 1 MGPLPRTVELFDVLSYVWGLFEILCRYQNIW-NINLQLRPSLITGIMKDSGNKPPGLL 59  
Db 1 MPKLR-IDFYEDVISPYVIAFEVFOKLETQWKGVTIRYIPFFLGAVMKESGNRRPPAML 59  
QY 60 PRKGLYMANDLKLRLHQLPIHPKDFLSVNLKSGLSAMRFLTAVNLEHPEMLEKASR 119  
Db 60 PARSIMMTDLKRTAKFWIDIPLTPPPLFMEWIKRYTTTGAMKVLVLVQSDKELMLRAAR 119  
QY 120 ELWMRVWSRNEDITEPQSIILAAAEKAGMSAQGLLEKIATPKVNQNLKETTEAAACRYG 179  
Db 120 EMVRLWSSEKIFEDQDFVEVLKAVGV--KNPEQIVKSKDEKIKILMENTNKGVDLM 177  
QY 180 AFGLP-ITVAHVDDQTHMFGSDRMELLHLHLLGEKMGPIP 219  
Db 178 AVGAPWINVHTEGSEHSFFGSDRFLIADLLQOP--OPLP 216

RESULT 2  
T34201  
hypothetical protein D2024.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34201  
R:Du, Z.; Gattung, S.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid D2024.  
A:Reference number: Z21488

A:Accession: T34201  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-225 <DU2>  
A:Cross-references: EMBL:U41011; PIDN:AAA82289.1; CESP:D2024.7  
C:Genetics:  
A:Gene: CESP:D2024.7  
A:Introns: 51/1; 177/3

Query Match 25.2%; Score 298.5; DB 2; Length 225;  
Best Local Similarity 33.0%; Pred. No. 2.7e-18;  
Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;  
QY 3 PLPTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGL---L 59  
Db 2 PNRKVVKEFFDVSPYSYGFEGITRHSVWKPIQMKPFFAGVVRHTEN--PGLPLRI 59  
QY 60 PRKGLYMANDLKLRLHHLQIPIHFPPKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
Db 60 PIKERYMHKDLFLSAQYWGIPFRLPKDYTNMLNTSSIVPQILVASQLRDNVLMEDYAR 119  
QY 120 ELWMRVWSRNEI--TEPOSILAAAEKAGSAEQAGLLEKATPKVNKOLKETTAACRY 178  
Db 120 GLWHRFYAYGKPIFTKSKQ---VAEVLRLDGLHVKDVELVMMSDSAEVKNILRENTDEIGN 176  
QY 179 GAFGLP---ITVAHVDGQT--HMLFGSDRMELLALHLLGEKWMGPI 218  
Db 177 GCFGAPMHIITDGH--GKVLQIVFGSDRLPQVADFLAEPKGP 218

RESULT 3  
G83629  
hypothetical protein PA0118 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83629  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Llm,  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83629  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <STO>  
A:Cross-references: GB:AE004450; GB:AE004091; NID:g9945943; PIDN:AA03508.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0118  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

\*Query Match 16.7%; Score 198; DB 2; Length 195;  
Best Local Similarity 26.18; Pred. No. 9.9e-10;  
Matches 55; Conservative 41; Mismatches 93; Indels 22; Gaps 5;  
QY 4 LPTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGLPRKG 63  
Db 1 MSKQIEFFDFGSPPTTVAWTLQPLPRAAAGASIAWRPMLLGGVFRKATGNHSPLEVPAGK 60  
QY 64 LYMANDLKLRLHHLQIPI---HPPKDFLSVMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
Db 61 RYTLHLIARVAKRYGVPLAFNPAPINTLTLM-----RGAQGYLGG-----EGFQPYLK 109  
QY 120 ELWMRVWSRNEITEPOSILAAAEKAGSAEQAGLLEKATPKVNKOLKETTAACRYG 179  
Db 110 AVEALWVRQNLGKPEVVAQVLAEGADPDDE---FLRLVGDEQVKEGLKATTEEAARRG 166  
QY 180 AFGCLPITVAHVDGQTHMLFGSDRMELLALHL 210  
Db 167 VFGAPSFV---GDQLFFGQDRDLDFVAEVL 193

## RESULT 4

S72164  
2-hydroxychromene-2-carboxylate isomerase homolog - Rhizobium leguminosarum bv.  
C:Species: Rhizobium leguminosarum bv. viciae  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 29-Sep-1999  
C:Accession: S72164  
R:Brito, B.; Palacios, J.M.; Ruiz-Argueso, T.; Imperial, J.  
Biochim. Biophys. Acta 1308, 7-11, 1996  
A:Title: Identification of a gene for a chemoreceptor of the methyl-accepting ty  
A:Reference number: S72162; MUID:96328256  
A:Accession: S72164  
A:Molecule type: DNA  
A:Residues: 1-194 <BRI>  
A:Cross-references: EMBL:U23040; NID:g780654; PIDN:AA044312.1; PID:g780657  
A:Experimental source: strain UPW791  
C:Genetics:  
A:Genome: plasmid  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 13.9%; Score 164.5; DB 2; Length 194;  
Best Local Similarity 25.2%; Pred. No. 7.4e-07;  
Matches 54; Conservative 34; Mismatches 91; Indels 35; Gaps 8;  
QY 6 RTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPPGL-LPRKGL 64  
Db 2 RTLDFYDYRSPYSFLALSQVRK---MDVEIAFHPLIEIGLDMKQGVNVPVTSITCAPKGR 57  
QY 65 YMANDLKLRLHHLQIPIHFPPKDFLSV-----MLEKSGLSAMRFLTAVNLEHPEMLEK 116  
Db 58 YVMTDQIRWAVHYGSLNHPHQLLEIDASRLRLATLVAGQLGAMP--TAV----- 105  
QY 117 ASRELMMRVWSRNEITEPOSILAAAEKAGSAEQAGLLEKATPKVNKOLKETTAAAC 176  
Db 106 --EAFINAIWSAPAPLATAAAEVAALVGAAGLDAEE---LAERMDPEAAQDLDEATANAV 160  
QY 177 RYGAFLPITVAHVDGQTHMLFGSDRMELL-AHL 209  
Db 161 SRGVGAPTLFV---GDEMFFGNDRLHFMQGH 190

## RESULT 5

H87328  
conserved hypothetical protein CC0643 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: H87328  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidell  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Frae  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: H87328  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-206 <STO>  
A:Cross-references: GB:AE005673; NID:gl3421856; PIDN:AAK22628.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0643  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 12.48; Score 147.5; DB 2; Length 206;  
Best Local Similarity 24.0%; Pred. No. 2.3e-05;  
Matches 50; Conservative 43; Mismatches 86; Indels 29; Gaps 8;

QY 6 RTVELFYDVLSPYSWLGFEILCRYQNIWNLQRLPSLITGIMKDSGNKPP-----GLLP 60  
Db 8 KSIDFIDFGSPNAYLSWKLLPEIAAROGAARVRLPCLLGGIFKATGNQAPMTAFGGVK 67



RESULT 7  
T31286  
2-nitrotoluene dioxygenase (EC 1.14.-.-) Rieske iron-sulfur component - Sphingomonas aromaticivorans  
C:Species: Sphingomonas aromaticivorans  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T31286  
A:R:Romane, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.; G  
A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromatic  
A:Reference number: Z20992  
A:Accession: T31286  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <ROM>  
A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378427; PID:AA04010.1  
C:Genetics:  
C:Gene: nahD

## STATISTICS:

7



```
QY 116 KASRELWVRVSRNEDITEPOSTILAAAEKAGM-----SAEQAGGLEKIAATPKVKNLK 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 RUPGGYIEIVERGTTLSAGERQLIARAVLFDKIFILDEATSNVDVTETRIQEALE 522
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 170 ETE 173
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 523 ELSK 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
C69127
flavoprotein A homolog (II) - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
C:Accession: C69127
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Oiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.;
Kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514
A:Accession: C69127
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <MTH>
A:Cross-references: GB:AE000809; GB:AE000666; NID:G2621265; PIDN:AAB84726.1; PID:G262126
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: WPH220
A:Start codon: TTG
C:Superfamily: Methanobacterium flavoprotein A
C:Keywords: flavoprotein

Query Match 7.6%; Score 90.5; DB 2; Length 410;
Best Local Similarity 25.1%; Pred. No. 4.5;
Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;

QY 13 DVLSPSYWLGFELC-----RYQNIWINLQRLPSLTGIMKDSGNK----- 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 EVLKRY---GSEIICTAKAAAGLQHSYIPQDTPMQ---TVKGTGSDTLGGKTLTFLEAP 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 55 ----PPG---LLPRKGLYMANDKLLRHHLQIPHPKDFLSVNLKGLSLSAMRF-----L 103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 MLHPDMSFTLEEGILFSND--AFQHLICISKREDKDVPEAVLMD--AAKMFYANLL 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 TAVNLEHPEMLEKAS--RELWM-----RVWSRNEDITEPOSILAAAEKAGMSA 149
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 TPLS---PLVLRKFSEVKELGLEKIGMTAPSHGQIW-----TEPLKIITAA-----YT 252
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 EQAOGLEKIAATPKVKNLKETEACRYGAGFLPITVAHVDGTHMFGSDRMELLAHL 209
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 DWATCKCRDKAT-IYDTHYSTRMLAHAAEGL--MAADVDSMHFLHEDERSEIVKNI 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 LCEK-----WMGPPI 219
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 LESKAVFIGSPWFNGPFP 328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
D95899
probable dipeptide ABC transporter permease and ATP-binding protein SMB20478 [imported]
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: D95899
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95899
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <KUR>
```

```
A:Cross-references: GB:AL591985; PIDN:CAC48860.1; PID:gl5140333; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-H
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher,
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Le
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Y
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMB20478
A:Genome: plasmid

Query Match 7.6%; Score 90.5; DB 2; Length 600;
Best Local Similarity 21.6%; Pred. No. 7.4;
Matches 33; Conservative 27; Mismatches 44; Indels 49; Gaps 6;

QY 34 NINLQRLPSLTGIMKDSGN-----KPPGLLPKRGLYMANDLKLRRHHLQI 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 DVSLHVKGCECLGLIGESGSKSVTALSVMGLVASPPGVIRNGAVYLGND----- 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 PIHFPKDFLS-----VMLEKGLSLSAMRFITAVNLEHPEMLEKASRELWVRVSRNEDITE 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 -----DVLSPMPETRLIAKRGSLAYVFQDPLTLHP--MYPVGQV-----EEATAA 435
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 POSILAAAEKAGMSAFOAQOGLLEKIAATPKVKNO 167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 HQVRAAERR-----EKAVALLEKVGIPDARER 463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
SNECPI
pitriylsin (EC 3.4.24.55) precursor [invalidated] - Escherichia coli
A:Alternate names: endopeptidase Pi; proteinase III
C:Species: Escherichia coli
C:Date: 31-Mar-1993 #sequence_revision 31-Oct-1997 #text_change 18-Feb-2000
C:Accession: F65064; A29093; A25765; B25532
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley,
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: F65064
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-962 <BLAT>
A:Cross-references: GB:AE000365; GB:U00096; NID:G2367163; PIDN:AAC75860.1; PID:G47391
A:Experimental source: strain K-12, substrain MG1655
R:Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.
Gene 54, 185-195, 1987
A:Title: Analysis of the regulatory region of the protease III (ptr) gene of Esc
A:Reference number: A29093; MUID:88005781
A:Accession: A29093
A:Molecule type: DNA
A:Residues: 1-276, 'W', 285-296 <CLA>
A:Cross-references: GB:M17095; NID:gl47390; PIDN:AAA24436.1; PID:gl47391
A:Experimental source: strain K12
A:Note: part of this sequence, including the amino end of the mature protein, was
R:Finch, P.W.; Willson, R.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.
Nucleic Acids Res. 14, 7695-7703, 1986
A:Title: Complete nucleotide sequence of the Escherichia coli ptr gene encoding I
A:Reference number: A25765; MUID:87040734
A:Accession: A25765
A:Molecule type: DNA
A:Residues: 1-962 <FIN>
A:Cross-references: GB:X06227; NID:G42560; PIDN:CAA29576.1; PID:G42561
R:Becker, A.B.; Roth, R.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 3835-3839, 1992
A:Title: An unusual active site identified in a family of zinc metalloendopeptid
A:Reference number: A38854; MUID:92237263
```

A:Contents: annotation; active site  
C:Genetics:  
A:Gene: ptr  
A:Map position: 61  
C:Function:  
A:Description: endopeptidase degrades small peptides [validated, MUID:92237263]  
A:Pathway: protein degradation  
C:Superfamily: insulysin  
C:Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein degradation  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-962/Product: pitrilysin #status experimental <MAT>  
F:88,92/Binding site: zinc (His) #status experimental  
F:91/Active site: Glu #status experimental

Query Match 7.6%; Score 89.5; DB 1; Length 962;  
Best Local Similarity 26.3%; Pred. No. 17;  
Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;  
QY 126 WSRNEDITEPQSILAAAEKAGMSAEQA-----OGLLEKIATPKVKNQ 167  
Db 731 WCRNKDVVDKQSVIFEKAGNSTDSALAAVFVPTGYDEVTSSAYSLLGQIVQPFYNQ 790  
QY 168 LKETTAAACRYGAFGLPTTVAHVQDQTHMFGSDRMELLALLGEXWGPPIPPA 221  
Db 791 LR--TEEQLGAYVAFPMVSGRQWGMGFLQSNKQ---PSFLWERYKAPFFPTA 839

Search completed: May 20, 2002, 08:39:58  
Job time: 315 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: May 20, 2002, 08:39:33 ; Search time 26.65 Seconds  
(without alignments)  
1467.049 Million cell updates/sec

Title: US-09-441-723-1  
Perfect score: 1185  
Sequence: 1 MGPLPRTVELFDVLSVPSW.....AHLLEKWMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL19:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_virus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	888	74.9	226	11	Q9DCM2	Q9DCM2	mus musculus
2	198	16.7	195	16	Q9I714	Q9I714	pseudomonas
3	195.5	16.5	207	2	Q93NA9	Q93NA9	burkholderi
4	164.5	13.9	194	2	Q52782	Q52782	rhizobium 1
5	149	12.6	195	2	Q9X9Q7	Q9X9Q7	spingomona
6	147.5	12.4	206	16	Q9AAF8	Q9AAF8	caulobacter
7	145.5	12.3	312	5	Q22312	Q22312	caenorhabdi
8	140	11.8	197	2	Q85994	Q85994	spingomona
9	135	11.4	196	2	Q9ZHH4	Q9ZHH4	burkholderi
10	114	9.6	196	2	Q9WXH0	Q9WXH0	alcaligenes
11	114	9.6	253	16	Q9A923	Q9A923	caulobacter
12	113	9.5	241	2	Q93HB9	Q93HB9	streptomyces
13	106	8.9	199	2	Q923X5	Q923X5	raistonia s
14	105	8.9	196	2	Q93CP1	Q93CP1	burkholderi
15	97.5	8.2	238	2	Q9Z167	Q9Z167	pseudomonas
16	91	7.7	581	16	Q9WXQ0	Q9WXQ0	thermotoga

17	90.5	7.6	410	17	O26322	O26322	methanother
18	90.5	7.6	199	16	Q92W85	Q92W85	rhizobium m
19	89	7.5	199	2	Q51499	Q51499	pseudomonas
20	87	7.3	926	16	Q911V3	Q911V3	pseudomonas
21	85	7.2	323	2	Q93H27	Q93H27	atryptomyce
22	83.5	7.0	308	2	Q92NM7	Q92NM7	exiguobacte
23	83.5	7.0	1007	10	Q92VD4	Q92VD4	arabidopsis
24	83	7.0	863	12	Q93124	Q93124	human calic
25	82.5	7.0	263	2	O87197	O87197	thermus the
26	82	6.9	203	2	O07298	O07298	pseudomonas
27	82	6.9	296	2	O52799	O52799	rhizobium 1
28	81.5	6.9	206	2	Q93CM5	Q93CM5	pseudomonas
29	80.5	6.8	413	10	Q9LUP2	Q9LUP2	arabidopsis
30	80	6.8	221	5	O09355	O09355	nosema locu
31	80	6.8	402	17	O26282	O26282	methanother
32	80	6.8	691	5	Q95R88	Q95R88	drosophila
33	80	6.8	704	5	Q9VVT8	Q9VVT8	drosophila
34	80	6.8	821	2	O30926	O30926	escherichia
35	79	6.7	230	16	Q9RSB4	Q9RSB4	deinococcus
36	78.5	6.6	223	17	Q973S1	Q973S1	sulfolobus
37	78.5	6.6	333	16	Q9HYF7	Q9HYF7	pseudomonas
38	78.5	6.6	493	10	Q9SVJ3	Q9SVJ3	arabidopsis
39	78.5	6.6	693	5	Q9VS91	Q9VS91	drosophila
40	78.5	6.6	778	16	Q92VY3	Q92VY3	rhizobium m
41	78.5	6.6	1072	6	O77690	O77690	bos taurus
42	78	6.6	248	17	O29435	O29435	archaeoglob
43	78	6.6	313	17	Q9YF18	Q9YF18	aeropyrum p
44	78	6.6	338	16	O84200	O84200	chlamydia t
45	78	6.6	687	11	Q9EPX1	Q9EPX1	mus musculu

ALIGNMENTS

RESULT	1
Q9DCM2	PRELIMINARY; PRT; 226 AA.
AC	Q9DCM2;
DT	01-JUN-2001 (TREMREL. 17, Created)
DT	01-JUN-2001 (TREMREL. 17, Last sequence update)
DT	01-JUN-2001 (TREMREL. 17, Last annotation update)
DE	O610025119RIK PROTEIN.
GN	O610025119RIK.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N A.
RC	STRALN=C57BL/6J; TISSUE=KIDNEY;
RX	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA	Schirali L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez L., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA	Hayashizaki Y.,
RA	Functional annotation of a full-length mouse cDNA collection.*;
RT	Nature 409:685-690(2001).
RL	EMBL; AK002661; BAB22268.1; -
DR	MGI:1923513; O610025119RIK.



Db 162 AYQGVGAP--MMFVDDQ--IFWGNDRDLFESYL 193

RESULT 6

Q9AAAF8 PRELIMINARY; PRT; 206 AA.

AC Q9AAAF8;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE HYPOTHETICAL PROTEIN CC0643.

GN CC0643.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter

OX NCBI\_TaxID=69394;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=ATCC 19089 / CB15;

RC MEDLINE=21173698; PubMed=11259647;

RA Nierman W.C., Feldblyum T.V., Laub M.R.K., Ohta N., Maddock J.R.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Stephens C., Phadke N.D., Ely B.,

RA DeBoy R.T., Dodson R.J., Newton A., Swinn M.L., Haft D.H.,

RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005740; AAK22628.1; -

DR TIGR; CC0643; -

KW Hypothetical protein; Complete proteome.

SEQUENCE 206 AA; 22058 MW; 634328E44B88489A CRC64;

Query Match 12.4%; Score 147.5; DB 16; Length 206;

Best Local Similarity 24.0%; Pred. No. 5.8e-06;

Matches 50; Conservative 43; Mismatches 86; Indels 29; Gaps 8;

QY 6 RTVELFYDVLSPYSWLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPP-----GLLP 60

Db 8 KSIDIFDGSFNAYLSWKLPEAARAGAAVRLICLLGGIFKATGQAQPIAFGGVKG 67

QY 61 RGLYMANDLKLRLH-----LQIPHPKDFLSVWLEKSLSAMRFLTAVNLEHPMLEK 116

Db 68 KMDYEMLETRFRFAHGLTAFRNPFPVN--TLLMRGMIAAQIGVA-----E 115

QY 117 ASRELWNR-VWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKETEAA 175

Db 116 AYLEAMKGMWEDGLKDDPEVFVATANAAGL---DGAALLAATGDAEVKAEVLVANTEAA 172

QY 176 CRYGAFGLPITVAHVGDQTHMLFGSDRM 203

Db 173 VARGTEGIPTFFV--GE-EIFFGKERL 196

RESULT 7

Q22312 PRELIMINARY; PRT; 312 AA.

AC Q22312;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE T07E3.3 PROTEIN.

GN T07E3.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=BRISTOL N2;

RC Wilton R., Alnscough R., Anderson K., Baynes C., Berks M.,

UPW791.";

RL Blochim. Biophys. Acta 1308:7-11(1996).

DR EMBL; U23040; AAC44312.1; -

DR InterPro; IPR004287; HCCA\_isomerase.

DR Pfam; PF03046; HCCA\_isomerase; 1.

KW Isomerase.

SEQUENCE 194 AA; 21153 MW; 0421C131B37BA435 CRC64;

Query Match 13.9%; Score 164.5; DB 2; Length 194;

Best Local Similarity 25.2%; Pred. No. 1.4e-07;

Matches 54; Conservative 34; Mismatches 91; Indels 35; Gaps 8;

QY 6 RTVELFYDVLSPYSWLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPPGL-LPRKGL 64

Db 2 RILDFEYDYSYFSLALSQVRK---MDVEIAFHPLEIGDLMKQGVNPTSITCAPKR 57

QY 65 YMANDLKLRLHLLQIPIHPKDFLSV-----MLEKGSLSAMRFLTAVNLEHPMLEK 116

Db 58 YVMTDIQRWAVHYGVSLSNHPOLLEIDASRLRLATLVAGQGLAMP--TAV----- 105

QY 117 ASRELWNRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKETEAA 176

Db 106 --PAIWAISAPATAAEVAVLGAAGLDAEE---LAERMDPEAQAQDLLEATANAV 160

QY 177 RYGAFLPITVAHVGDQTHMLFGSDRMELL-AHL 209

Db 161 SRGVFGAPTFLV-----GDMEFFGNDRLHFMQGH 190

RESULT 5

Q9X9Q7 PRELIMINARY; PRT; 195 AA.

AC Q9X9Q7;

DT 01-NOV-1999 (Tremblrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE.

GN NSAD.

OS Sphingomonas sp.

OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;

OC Sphingomonas.

OX NCBI\_TaxID=28214;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=BN6;

RA Keck A.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; U65001; AAD45416.1; -

DR InterPro; IPR004287; HCCA\_isomerase.

DR Pfam; PF03046; HCCA\_isomerase; 1.

KW Isomerase.

SEQUENCE 195 AA; 22064 MW; 7E6FBAC96B3C6316 CRC64;

Query Match 12.6%; Score 149; DB 2; Length 195;

Best Local Similarity 23.1%; Pred. No. 3.9e-06;

Matches 50; Conservative 43; Mismatches 91; Indels 32; Gaps 7;

QY 4 LPRTVELFYDVLSPYSWLGFEILCRYQNIWINLQRLPSLITGIMKDSGNKPPG---LLP 60

Db 1 MTKTIDFYDFISPSYLAQVKLPDLARRTGCVIEYRPIDPEAKIAGNYPGNSREYVP 60

QY 61 RGLYMANDLKLRLHLLQIPIHPKDF-----LSVMLEKGSLSAMRFLTAVNLEHPML 114

Db 61 KIKVMAA-DLERWAAYEVLTPFASPACSDNWCNCAALYARQDAEAVTAA----- 111

QY 115 EKASRELWNRVWSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKETEAA 174

Db 112 -----YIRIWIGIDIPRONELRGCAEDVGLDAD---ALCEFVRSAGGGEYRKARTO 161

QY 175 ACRYGAFGLPITVAHVGDQTHMLFGSDRMELLAHL 210





```

RESULT 10
39WXHO PRELIMINARY; PRT; 196 AA.
ID Q9WXH0
AC Q9WXH0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DD 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ISOMERASE.
GN PHND
OS Alcaligenes faecalis
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Alcaligenes.
OX NCBI_TaxID=511;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AFK2;
RA Kiyohara H., Tabata Y., Takizawa N.;
RT "A phenanthrene degradative gene cluster in Alcaligenes faecalis
RT AFK2.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB024945; BAA76325.1; -.
DR InterPro; IPR004287; HCCA_isomerase.
DR Pfam; PF03046; HCCA_isomerase; 1.
DR SEQUENCE 196 AA; 22355 MW; C02757978BEE2E2 CRC64;
SQ
Query Match 9.6%; Score 114; DB 2; Length 196;
Best Local similarity 21.1%; Pred. NO. 0.0077;
Matches 44; Conservative 40; Mismatches 99; Indels 26; Gaps 7;

QY 8 VELFYDLSPYSMLGF---EILCRYQINWINQLRPSLTIGMKDSGNKPPGL--LPR 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 LSFYDFVSFYSYLASRLPEIVORY---GISVSYKPIDIACAKRAIGNVGPNSRDMPV 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 KGLYMANDLLRLHHLQIPIHPKDFLSVMLEKGLSAMRELTAVNLEHPMELEKASREL 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 58 KLTHLSRDLQWAQRYGTPKLPFPSPDSRRLNTGF----FYAAGEAREAEYVRR-- 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 WMRYWSRNEDITEPQSILAAAEKAGMAEQAGLLEKATPKVKNQKLTETAAACRYGAF 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 FHLTWGCGQAFSGEPVLRSIASEMGWNVDDFMQFTDSV---DGANEYKQSIDEGIARSVF 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 182 GLPITVAHVDDGQTHMLFGSDRMELLALHL 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 166 GVPWVI----GDEMWMGNDRDLDFVDEYL 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
Q9A923 PRELIMINARY; PRT; 253 AA.
ID Q9A923
AC Q9A923;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DD 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE DEHYDROGENASE, PUTATIVE.
GN CC1171.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 19089 / CB15;
RX MEDLINE=211173698; PubMed=11259647;
RA Eisenman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
```

QY 217 -PIPPA 221  
 Db 225 LPVPA 230

RESULT 13  
 Q923X5  
 ID Q923X5 PRELIMINARY; PRT; 199 AA.  
 AC Q923X5;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE 2-HYDROXYCHROMENE CARBOXYLATE ISOMERASE.  
 GN NAGD.  
 OS Ralstonia sp. U2.  
 OG Plasmid pMWU2.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=70356;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=U2;  
 RX MEDLINE=98233751; PubMed=9573207;  
 RA Fuenmayor S.L., Wild M., Boyes A.L., Williams P.A.;  
 RT "A gene cluster encoding steps in conversion of naphthalene to  
 gentisate in Pseudomonas sp. strain U2.";  
 RL J. Bacteriol. 180:2522-2530(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=U2;  
 RX MEDLINE=20576173; PubMed=11133965;  
 RA Zhou N.Y., Fuenmayor S.L., Williams P.A.;  
 RT "nag genes of ralstonia (formerly pseudomonas) sp. Strain U2 encoding  
 enzymes for gentisate catabolism.";  
 RL J. Bacteriol. 183:700-708(2001).  
 DR EMBL; AF036940; AAD12617.1;  
 DR InterPro; IPR004287; HCCA\_isomerase.  
 DR Pfam; PF03046; HCCA\_isomerase; 1.  
 KW Isomerase; Plasmid.  
 SQ SEQUENCE 199 AA; 23067 MW; 219466E06AC04E47 CRC64;

Query Match 8.9%; Score 106; DB 2; Length 199;  
 Best Local Similarity 24.9%; Pred. No. 0.044;  
 Matches 52; Conservative 32; Mismatches 91; Indels 34; Gaps 8;

QY 8 VELFYDVLSPYSWLGFEILCRQNTQNIWNLQRLPSLITGIMKDSGNKPPGL--LPRKGLY 65  
 Db 5 VDFYDFLSFYSYLANHRLSVLAGRYGFSIQYHAIDLAKATAIGNIGPSNRDLKVLKY 64

QY 66 MANDKLLRHHLQIPIHFPKDLSVMLEKGSLSAMRFLTAVNLEHPMELEKASR----- 119  
 Db 65 LKVDLQWADLYRPLVFPNPSNRVAG-----LYYPAAREAAEYVRLVF 112

QY 120 -ELMWRVSRNEDITEPOSILA-AAEAKGMSAEQAQGLLEKIATPKVKNLKETEACR 177  
 Db 113 DSAWGKGWALDAD-----SLLAEVCDKLNWDLGERFDELSENAAKAYD---ETQQAID 164

QY 178 YGAFGLPITVAHVGDQTHMLFGSDRMELL 206  
 Db 165 RKVFGVP-TVFWDD---QMWNGNDRLMFL 189

RESULT 14  
 Q93CPI  
 ID Q93CPI PRELIMINARY; PRT; 196 AA.  
 AC Q93CPI;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE PUTATIVE ISOMERASE.  
 OS Burkholderia sp. DBT1.

OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC Burkholderia.  
 OX NCBI\_TaxID=161152;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBT1;  
 RA Di Gregorio S., Zocca C., Vallini G.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF404408; AAK96187.1;  
 KW Isomerase.  
 SQ SEQUENCE 196 AA; 22515 MW; CB5B111CEA6C08E1 CRC64;

Query Match 8.9%; Score 105; DB 2; Length 196;  
 Best Local Similarity 20.5%; Pred. No. 0.054;  
 Matches 42; Conservative 40; Mismatches 105; Indels 18; Gaps 5;

QY 4 LPTVELFYDVLSPYSWLGFEILCRQNTQNIWNLQRLPSLITGIMKDSGNKPPG--LLPR 61  
 Db 1 MTKNLFYDFISFYSYLANHRLSVLAGRYGFSIQYHAIDLAKATAIGNIGPSNRVQS 60

QY 62 KGLYMANDKLLRHHLQIPIHFPKDLSVMLEKGSLSAMRFLTAVNLEHPMELEKASREL 121  
 Db 61 KMKVLLADMRWAARYDVPLTFPGKFCGERWNIGVLYA-----DKKRDTEAYVTET 111

QY 122 WWRVSRNEDITEPOSILA-AAEAKGMSAEQAQGLLEKIATPKVKNLKETEACRYGAF 181  
 Db 112 YHRWGLGINPSDEGOLGEVAEKMGWIAE---FLTFVSSPEGQATFRKSCVEAHARGVF 168

QY 182 GLPITVAHVGDQTHMLFGSDRMELL 206  
 Db 169 GAPIMV---GE-EVWNGNDRLMFL 189

RESULT 15  
 Q92167  
 ID Q92167 PRELIMINARY; PRT; 238 AA.  
 AC Q92167;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE 2-HYDROXYCHROMENE-2-CARBOXYLATE DEHYDROGENASE.  
 GN NAHD.  
 OS Pseudomonas stutzeri (Pseudomonas perfectomarina).  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=316;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AN10;  
 RX MEDLINE=99365311; PubMed=10433976;  
 RA Bosch R., Garcia-Valdes E., Moore E.R.B.;  
 RT "Genetic characterization and evolutionary implications of a  
 chromosomally encoded naphthalene-degradation upper pathway from  
 Pseudomonas stutzeri AN10.";  
 RL Gene 236:149-157(1999).  
 DR EMBL; AF039533; AAD02142.1;  
 DR InterPro; IPR004287; HCCA\_isomerase.  
 DR Pfam; PF03046; HCCA\_isomerase; 1.  
 SQ SEQUENCE 238 AA; 26359 MW; 29B972C1CCF7C71B CRC64;

Query Match 8.2%; Score 97.5; DB 2; Length 238;  
 Best Local Similarity 23.5%; Pred. No. 0.35;  
 Matches 53; Conservative 37; Mismatches 103; Indels 33; Gaps 9;

QY 8 VELFYDVLSPYSWLGFEILCRQNTQNIWNLQRLPSLITGIMKDSGNKPPGL--LPRKGLY 65  
 Db 19 VDFYDFLSFYSYLANHRLSVLAGRYGFSIQYHAIDLAKATAIGNIGPSNRDLKVLKY 78

QY 66 MANDKLLRHHLQIPIHFPKDLSVMLEKGSLSAMRFLTAVNLEHPMELEKASRELMMR 124  
 Db 79 LMVDLKRWAELVGLPFLFPANYSQRMNAGLYSGAETQTAAYV-----NTVFNA 128



